National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

University of Alberta

A study of *P* element alleles of the *vestigial* locus in *Drosophila melanogaster*

by

Brian Ernest Staveley ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirement for the degree of Doctor of Philosopy

Department of Genetics

Edmonton, Alberta
Fall, 1995

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-612-06294-5

Canada

University of Alberta
Edmonton

Canada T6G 2E9

Department of Biological Sciences
Faculty of Science

CW 405 Biological Sciences Centre
Telephone: Office (403) 492-3308
Fax (403) 492-9234
Chair (403) 492-0672

June 27, 1995

To Whom it may Concern:

Brian Staveley has just recently defended his Ph.D. thesis. Two of the chapters are closely related to his published papers. I am a co-author on both of these papers, and on behalf of myself and the other co-authors (whom I have consulted), I give permission to Staveley to use a version of these publications in his thesis. The actual citations for the two papers are as follows:

① Staveley, B.E., S. O'Keefe, R.B. Hodgetts and J.B. Bell (1994) Targetting of an enhancer trap to vestigial. Develop. Biol. 165: 290-293.

② Staveley, B.E., T.R. Heslip, R.B. Hodgetts and J.B. Bell (1995) Protected P-element termini suggest a role for IRBP in transposase-induced gap repair in Drosophila melanogaster. Genetics 139: 1321-1329.

Please telephone me at 5382 if you require further information on this matter.

Sincerely,

John B Bell

Attention Brian Staveley

6/27/95

Dr. Tim R. Heslip
Developmental Biology Center
Univ. of California, Irvine
Irvine, CA 92717

To Whom it May Concern,

I, Tim Heslip, as a co-author of the published paper "Protected P-Element Termini Suggest a Role for Inverted-Repeat-Binding Protein in Transposase-Induced Gap Repair in Drosophila melanogaster" do hereby give permission to Brian E. Staveley to include a version of said paper in his Ph.D. thesis.

Sincerely

Tim Heslip

University of Alberta

Library Release Form

Name of Author: Brian Ernest Staveley

Title of Thesis: A study of *P* element alleles of the *vestigial* locus in *Drosophila melanogaster*

Degree: Doctor of Philosophy

Year this degree granted: 1995

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Brian Staveley

R.R. #4,
Orangeville, Ontario, Canada

June 26, 1995

Occasionally flies appear without wing, but this character is not inherited as rule, and is due to some difficulty in unfolding the primordia of the wings. But in some of the stocks of truncated wings I have obtained a considerable number of flies with tiny scales in place of wings. In one culture there appeared 11 flies with scales, instead of wings, amongst 125 winged flies. Although this stock is very sterile it seems not improbable that, in time, a wingless fly can be produced.

Morgan, T.H. 1911. Science **33**: 496-499.

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *A study of P element alleles of the vestigial locus in Drosophila melanogaster* by BRIAN ERNEST STAVELEY in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

John B Bell

JOHN B. BELL

ROSS B. HODGETTS

John Locke

JOHN LOCKE

MICHAEL A. RUSSELL

Charlotte Spencer

CHARLOTTE A. SPENCER

MA Russell

for DONAL A. HICKEY

June 30, 1995

This thesis is dedicated to my wife, Helene Schinko-Staveley,
for all the obvious reasons plus some.

ABSTRACT

The properties and interactions of the *vestigial* gene and various forms of an adjacent *P* element have been explored in this series of experiments through the isolation and characterization of derivatives of *vestigial*^{21} and by the continued examination of *vestigial* expression. The *vestigial* gene of *Drosophila melanogaster* is among the earliest identified and is pivotal to the formation of the Drosophila wing. The absence of *vestigial* function results in cell death due to an inability to establish positional cues during the patterning of the developing wing. A *P* element allele, *vestigial*^{21}, was generated in previous experiments to allow the molecular characterization of the *vestigial* gene. In the current investigations, an enhancer trap was targeted to the insertion site of *vestigial*^{21} to analyze the mechanisms that control *vestigial* transcription. The reporter gene expression pattern of the targeted line, *vestigial*^{lacZ1}, differs from that of *vestigial* and may respond to an organizing region in the centre of the wing imaginal disc. Examination of deletion derivatives of *vestigial*^{21} illustrates the inclination of some *P* elements to possess internal deletion breakpoints within the terminal inverted repeats. According to the synthesis-dependent strand annealing model, DNA synthesis is initiated at the two sides of a double-strand break left by *P* element excision. A template is copied and the strands anneal to repair the gap. The DNA sequences found at the internal breakpoints indicate that the inverted-repeat binding protein (IRBP) may protect the termini during repair. Analysis of the constitutions of the internally-deleted *P* elements and the resultant phenotypes suggests that *P* element transcriptional activity can influence the expression of *vestigial*. A *P* element-*vestigial* fusion transcript has been identified which appears to interfere with production of the *vestigial* transcript. In addition, examination of wild type *vestigial* transcription reveals that the *vestigial* pre-mRNA is subject to alternative splicing. In conclusion, this study contributes to the understanding of *P* element biology, furthers the study of the *vestigial* gene and examines the interactions between the gene and the transposon.

ACKNOWLEDGEMENTS

I would like to thank: my supervisor JOHN B. BELL who has been supportive throughout this study; ROSS B. HODGETTS for his collaborative efforts, temporary supervision in JOHN'S absence and his active participation in my committee; JOHN LOCKE for great committee service and thoughtful discussions; TOVE REECE for technical support and for holding down the fort; BILL CLARKE and GARY RITZEL for advice on a great number of techniques; CHARLOTTE SPENCER and DONAL HICKEY for participation in the thesis examination; GREG GLOOR for advice on *P* element biology; MICHAEL RUSSELL, HEATHER MCDERMID and JACK (R.C.) VON BORSTEL for inciteful discussions; TIM HESLIP for his collaboration with internally deleted *P* elements and providing the reversion data; SANDRA O'KEEFE for producing the PCR fragments of the insertion site of vg^{lacZ1}; KAY KOVITHIVONGS and MICHAEL VEEMAN for technical support;

And of course HELENE, for everything.

TABLE OF CONTENTS

Chapter	Title	page
1.	A brief review of the *vestigial* locus and the *P* element transposon	1
	Literature cited	15
2.	Targeting of an enhancer trap to *vestigial*	19
	Literature cited	31
3.	Protected *P* element termini in transposase-induced gap repair	33
	Literature cited	53
4.	A model of transcriptional interference for *P* element alleles of *vestigial*	56
	Literature cited	76
5.	Reexamination of *vestigial* expression	79
	Literature cited	97
6.	Insights into the *P* element and the *vestigial* gene	99
	Literature cited	116

LIST OF TABLES

CHAPTER

TABLE DESCRIPTION PAGE

CHAPTER 3

TABLE 3.1. Analysis of deletion breakpoints in derivatives of *vg*^{21} selected for strong *vestigial* phenotype. 39

TABLE 3.2. Analysis of deletion breakpoints in derivatives of *vg*^{21} selected for reversion. 41

TABLE 3.3. Analysis of deletion breakpoints in derivatives of *vg*^{21} selected for altered PCR fragment length. 43

CHAPTER 5

TABLE 5.1. Properties of the seven methionine codons of Vg, the *vestigial* protein. 88

TABLE 5.2. Properties of the seven methionine codons of a second conceptual *vestigial* protein (Vg-2). 91

LIST OF FIGURES

CHAPTER
FIGURE DESCRIPTION PAGE

CHAPTER 1

FIGURE 1.1. The molecular organization of the *vestigial* locus. 6

CHAPTER 2

FIGURE 2.1. Identification of a recruited enhancer trap at *vestigial* by Southern analysis. 24 & 25
FIGURE 2.2. ß-galactosidase staining of vg^{lacZ1} imaginal discs. 28

CHAPTER 3

FIGURE 3.1. Distribution of the breakpoint position in terminal internal deletions. 44
FIGURE 3.2. Terminal internal deletion derivatives of *vestigial*^{21}. 46
FIGURE 3.3. Possible role for IRBP in transposase-induced double-strand gap repair. 49

CHAPTER 4

FIGURE 4.1. The *P* element of *vestigial*^{21}. 61
FIGURE 4.2. Northern analysis of P element alleles of *vestigial*. 65
FIGURE 4.3. A possible *P* element fusion transcript at *vestigial*. 68
FIGURE 4.4. A transcription interference model of *vestigial*^{21} and its derivatives. 70

CHAPTER 5

FIGURE 5.1. The 5' region of the *vestigial* transcription unit. 85
FIGURE 5.2. The *vestigial* exon 1/exon 3 splice junction. 87
FIGURE 5.3. The amino acid sequence of *vestigial* conceptual protein-2 (Vg-2). 90
FIGURE 5.4. A comparison of a 60 amino acid region of Vg-2 to PSF. 92

Chapter 1

A brief review of the *vestigial* locus and the *P* element transposon

In *Drosophila melanogaster*, most adult structures develop from cells that are initially sequestered during embryogenesis and which develop into imaginal discs during the three larval instar stages prior to pupation (for review see COHEN 1993). Eventually the discs become differentiated organs during metamorphosis. Throughout the progressive stages of development, the wing discs are sequentially subdivided into compartments through the establishment of three different boundaries: the anterior/posterior, the dorsal/ventral and the proximal/ distal axes (COHEN 1993; WILLIAMS and CARROLL 1993). The thorax has both ventral and dorsal appendages, the ventral legs and the dorsal wings and halteres. The genes involved in these processes can be classified as either essential to the development of both the ventral and dorsal thoracic imaginal discs or as required by only one of the two sets. The genes of the first category include *decapentaplegic*, *wingless* and *aristaless*. Development of the legs relies upon the activity of *Distal-less* which is not required in wing and haltere discs (COHEN and JURGENS 1989). In contrast, the *apterous*, *scalloped* and *vestigial* genes act in dorsal discs and not in leg discs (COHEN *et al*. 1992; CAMPBELL *et al*. 1992; WILLIAMS *et al*. 1993). Since mutations of these three genes are associated with defects of the wing and haltere but not in the formation of the thoracic body structures, it can be concluded that these genes are involved in the specification of the proximal/distal axis. The *vestigial* gene is the focus of the series of experiments presented in this thesis.

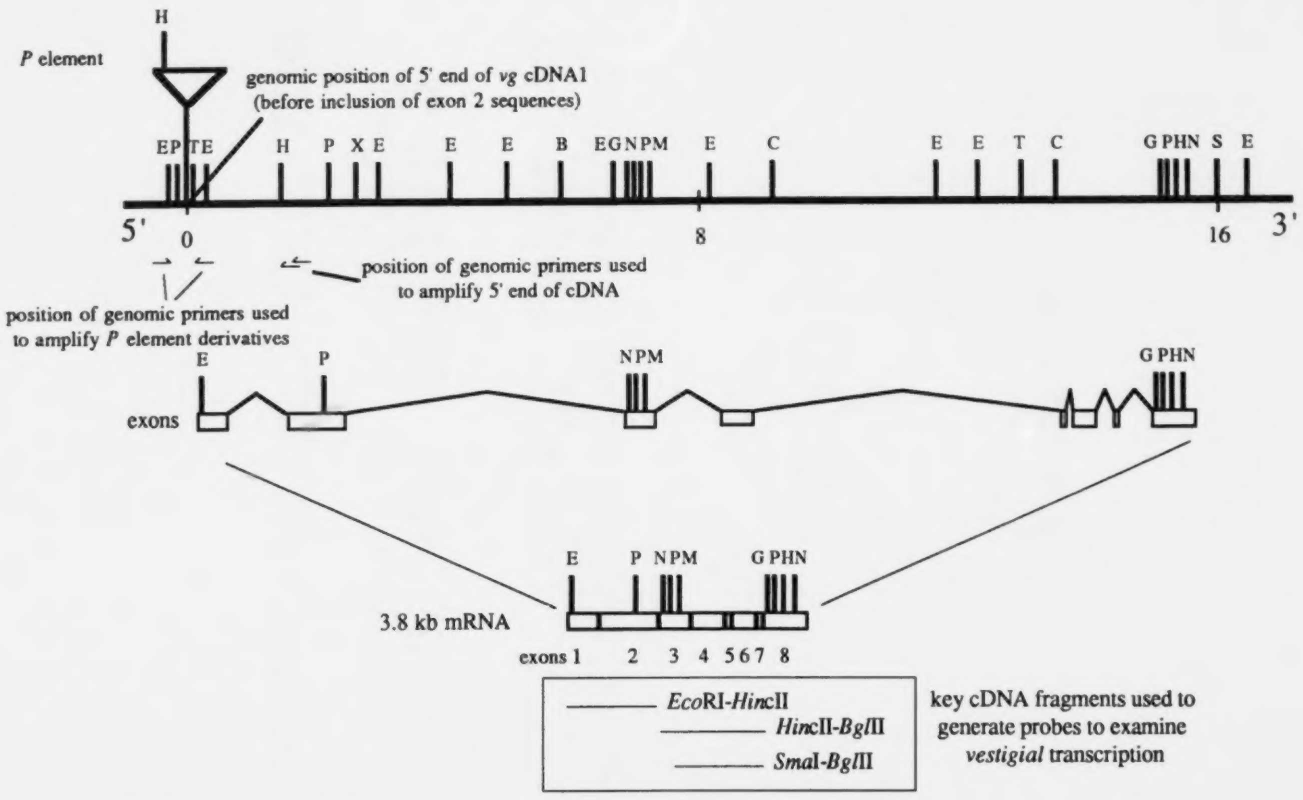
The *vestigial* gene: The *vestigial* mutant of *Drosophila melanogaster* is among the earliest isolated by the MORGAN group during the first years of Drosophila genetics (formerly named *wingless*: MORGAN 1911; BRIDGES and MORGAN 1919). The most obvious *vestigial* phenotype, the loss of wing structures, has been very well documented and a large number of *vestigial* alleles has been described (BRIDGES and MORGAN 1919; WILLIAMS 1989; LINDSLEY and ZIMM 1992). Strong *vestigial* alleles show a complete loss of wing margin structures and the wings are reduced to a small scale-like vestige of their normal size, hence the mutant designation. Less extreme alleles are common and show varying degrees of wing margin loss such as $vestigial^{strap}$, $vestigial^{antlered}$ and $vestigial^{nick}$ (BRIDGES and MORGAN 1919). The wing phenotypes of various alleles range from the cryptic, which appear to be wild type, through those displaying nicking or notching of the wing margin, to scalloped or strapped wings to the extreme classic *vestigial* wing. A cryptic allele is an allele which displays a wild type phenotype as a homozygote but in heterozygous combination with a strong allele fails to meet a threshold requirement of the gene's function to result in a mutant phenotype. The phenotypes of weak *vestigial* mutants are more extreme when heterozygous with a strong allele or hemizygous with a deficiency that spans the locus. In general, the extent of wing structure formation in these individuals is roughly intermediate between the classic *vestigial* phenotype and the hypomorph examined. This suggests that different threshold levels of the *vestigial* product are associated with specific amounts of wing material and/or discrete structures. The actions of the *vestigial* gene are also pleiotropic and include haltere reduction (analogous to the wing), erect postscutellar bristles, female sterility, low viability, pupal lethality, reduced adult size and developmental delay (LINDSLEY and ZIMM 1992). Although weak alleles (notch or strap wings) do not display defects in these traits, heterozygous combinations with strong alleles do. Since the amount of wing tissue can be directly correlated to a hypomorphic allelic series and defects in the other traits are only observed in alleles that display severe wing loss, it appears that different thresholds of the *vestigial* product are required for each trait such that formation of the wing margin requires the highest level to attain the wild type phenotype.

Heterozygotes carrying a strong *vestigial* allele and the wild type allele usually have normal wings but often these individuals display slight wing margin nicking (GREEN and OLIVER 1940). Some mutant genetic backgrounds, especially those that prolong development, increase the frequency of wing scalloping in the population of heterozygotes. The *vestigial* locus must be barely haplo-sufficient such that 50% of wild type *vestigial* product is usually enough to produce a complete wing. As well, a gradient must exist between 0% and 50% of normal *vestigial* activity which results in the formation of an increasing amount of wing tissue. The penetrance of $vestigial^{1}$ was examined in segmental aneuploids (GREEN 1946). The mutant phenotype was most common among $vestigial^{1}/vestigial^{1}/vestigial^{+}$ flies, less common in $vestigial^{1}/vestigial^{+}$ heterozygotes and very rare in $vestigial^{1}/vestigial^{+}/vestigial^{+}$ individuals. GREEN (1946) concluded that $vestigial^{1}$, in this light, is antimorphic. However, since most *vestigial* alleles, including $vestigial^{1}$, can be ordered into a descending allelic series with respect to wing structure and weak *vestigial* alleles in heterozygous combination with $vestigial^{1}$ display phenotypes that are intermediate when compared to homozygotes of the weak allele tested and $vestigial^{1}$, this allele is probably not an antimorph (WILLIAMS 1989). These findings suggest that *vestigial* regulates its own expression. The $vestigial^{83b27}$ mutation is a unique allele of *vestigial* and forms a second complementation group (ALEXANDROV and ALEXANDROVA 1987). This allele in heterozygous combination complements the wing and haltere effects of other *vestigial* alleles in a chromosome-pairing dependent manner (WILLIAMS *et al.* 1990a). Homozygous $vestigial^{83b27}$ mutants display extreme reductions in the wing and haltere but do not show scutellar or fertility defects. The *vestigial*^{83b27} mutants are wingless as homozygotes, however the wing vestiges have a morphology which is different from the other *vestigial* alleles. Instead of the tiny veined scale-like wing of *vestigial*^{1}, the mutant phenotype of *vestigial*^{83b27} ranges from no wing tissue to 'wisps' of wing material. The two dominant alleles of *vestigial*, *vestigial*^{W} and *vestigial*^{U}, both wingless as heterozygotes, result from chromosomal inversions that juxtapose *vestigial* and *invected* or *mastermind* respectively (WILLIAMS *et al.* 1990b). Reversion of the dominant *vestigial* phenotype of *vestigial*^{W} to reveal the dominant homeotic phenotype (KOVITHATHONGS and BELL, unpublished) suggests that the "wingless" phenotype of *vestigial*^{W} is caused by the misexpression of *vestigial* and not of *engrailed* or *invected.* An *apterous* cDNA under the control of the *vestigial* second intronic enhancer generates a reduction of wing tissue (WILLIAMS *et al.* 1994) and behaves in a manner similar to a dominant *vestigial* allele. The phenotype of *vestigial*^{U} may either be the result of misexpression of the *mastermind* or of a truncated version of the *vestigial* gene. Based upon the wide range of mutant phenotypes associated with defects of *vestigial*, the mechanisms controlling the *vestigial* gene must be complex.

Electron microscopy studies reveal that abnormal cell death occurs within the imaginal wing discs of *vestigial* mutant larvae (FRISTROM 1968; 1969). The dead cells undergo shrinkage and condensation followed by engulfment by a neighbouring cell. The degenerating cells are concentrated in the wing blade region of the imaginal disc. In a study to compare two different *vestigial* mutants, the pattern of cell death was shown to occur throughout the third larval instar in *vestigial*^{1} but only in the early third instar in the dominant *vestigial*^{U} allele (O'BROCHTA and BRYANT 1983). The cell death observed in *vestigial* mutants, which is concentrated in the presumptive posterior wing margin, does not result in duplications of the pattern elements of the wing disc as is often the case when cell death occurs in the wing (BOWNES and ROBERTS 1981). Since *vestigial* mutants cause cell death in the wing pouch which does not give rise to regulative growth, it is likely that the *vestigial* gene is directly involved in the establishment of the positional values in the developing wing.

The molecular characterization of the *vestigial* genomic region was initiated by the isolation of a *P* element insert allele of *vestigial*, *vestigial*^{21} (FIGURE 1.1; WILLIAMS and BELL 1988). A library was made from the genomic DNA of this strain, several *P* element containing clones were recovered and the corresponding wild type sequences

FIGURE 1.1. The molecular organization of the *vestigial* locus. A restriction map of the *vestigial* genomic region is presented (adapted from WILLIAMS and BELL, 1988). The insertion site of the *P* element of *vestigial*^{21} is indicated by the triangle above the map. The 5' extend of sequences present in *vg* cDNA1 (WILLIAMS *et al.*, 1990) is indicated. The position of primers used in the current study are displayed below the genomic map. The exon map of the full length 3.8 kb mRNA is aligned below the genomic sequences (WILLIAMS *et al.*, 1990). The restriction map of a full length cDNA and key cDNA fragments used in the preparation of probes are shown. The scale below the genomic map, in kilobases, is constant throughout.


H
P element
genomic position of 5' end of vg cDNA1
(before inclusion of exon 2 sequences)
EP TE
H
P X E
E
E
B
EGNPM
E
C
E
E
T
C
GPHN
S
E
5'
0
8
16
3'
position of genomic primers used
to amplify 5' end of cDNA
position of genomic primers used
to amplify P element derivatives
E
P
NPM
GPHN
exons
E
P NPM
GPHN
3.8 kb mRNA
exons 1 2 3 4 5 6 7 8
EcoRI-HincII
HincII-BglII
SmaI-BglII
key cDNA fragments used to
generate probes to examine
vestigial transcription

restriction sites: B: *Bam*HI; C: *Cla*I; G: *Bgl*II; E: *Eco*RI; H: *Hin*dIII; M: *Sma*I; N: *Hinc*II; P: *Pst*I; S: *Sal*I; T: *Sst*I; X: *Xho*I

were isolated. Lesions associated with a number of *vestigial* mutants were mapped to a 19 kb region defined as essential for *vestigial* function. Six independent cDNAs were isolated from an imaginal disc library by hybridization to probes of the *vestigial* genomic region (WILLIAMS *et al.* 1990a). The largest cDNA, *vg* cDNA1, is composed of a copy of the *vestigial* message plus additional sequences. Three isolates, *vg* cDNA2, *vg* cDNA3 and *vg* cDNA4, are nearly identical and slightly shorter than *vg* cDNA1. The fifth *vg* cDNA has only sequences from exons 1, 2, 3 and possibly 4. The unusual internally-deleted cDNA, *vg* cDNA6, has homology to the 5' and 3' ends of *vg* cDNA1 but seems to be missing sequences from the second *vestigial* exon. The genomic location of the exons correlates well with the lesions associated with *vestigial* alleles. A rare 3.8 kb transcript was found to be present in embryos older than 8 to 12 hours but was undetected in larval stages (WILLIAMS *et al.* 1990a). Interestingly, one cDNA probe which spans the majority of the *vestigial* exons, also recognized several smaller transcripts. The extra bands were discounted as artifacts due to cross hybridization and were not analyzed further. The smaller bands are consistent with the possibility that the *vestigial* transcript may be subject to alternative splicing.

Alternative splicing is important in the regulation of some genes in eukaryotes (MCKEOWN 1992; HODGES and BERNSTEIN 1994). A number of strategies to control gene activity can be accomplished by the production of different transcripts from a single precursor pre-mRNA. This process can be used to turn on and off gene functions and to generate alternate forms of a gene product having different functions. Alternative splicing may take many different approaches including the use of different promoters, alternative splice donor or acceptor sites, mutually exclusive exons and exon skipping. Many Drosophila genes have been demonstrated to undergo alternative splicing including the *P* element transcript (LASKI *et al.* 1986) and several genes in the sex determination pathway (BELL *et al.* 1988; MATTOX and BAKER 1991). Regulation of alternative splicing is often a reflection of tissue-specific or developmentally specific control strategies.

The *vestigial* message and protein are expressed in a predominant stripe along the dorsal/ventral boundary of the wing disc the site of the presumptive wing margin (WILLIAMS *et al.* 1991). The early activities of *wingless* and *apterous* establishes the border between the dorsal and the ventral compartments of the presumptive wing (WILLIAMS *et al.* 1993). Once formed, signals present at the border activate expression of the *vestigial* gene in a narrow band of cells spanning the dorsal/ventral boundary by inducing a regulatory element within the *vestigial* gene (WILLIAMS *et al.* 1994). The genetic properties of the *vestigial*^{83b27} mutant (ALEXANDROV and ALEXANDROVA 1987), which is deleted for sequences only within the second intron, indicate that control element(s) exist within the second intron (WILLIAMS *et al.* 1990a). Analysis of this mutant demonstrated that early wing disc expression is absent (WILLIAMS *et al.* 1991). A comparison of the *vestigial* sequences from *Drosophila melanogaster* and *Drosophila virilis* revealed a region of homology within the *vestigial* second intron which is conserved for 135 of 137 base pairs. A 750 base pair *Eco*RI fragment bearing the conserved region was cloned into a *lacZ* reporter gene construct and yielded stable transformed lines which displayed a ß-galactosidase expression restricted to a stripe of cells spanning the dorsal/ventral boundary in the wing and haltere discs. This suggests that the intronic sequences direct the expression of *vestigial* along the presumptive wing margin.

To investigate the activation of the enhancer, *apterous* clones were made in the presence of this reporter construct. In the ventral compartment, *apterous* clones revealed the reporter gene pattern of expression as seen in a wildtype background. Similar clones in the dorsal compartment resulted in *lacZ* activity at the clonal border in the cells which had lost *apterous* expression. This is consistent with *apterous* expression as a major determinant of dorsal fate (WILLIAMS *et al.* 1993) and its loss leads to a ventral-like fate and that where dorsal and ventral cells meet, the *vestigial* enhancer is induced. In *vestigial* wing discs, the early third instar discs appear normal but the wing pouch does not develop. The *scalloped* and *vestigial* proteins are the two early markers of the wing region in early discs (WILLIAMS *et al.* 1993). The reporter construct was introduced to *vestigial* and *scalloped* pupal-lethal backgrounds (WILLIAMS *et al.* 1994). ß-galactosidase expression in the late second instar to early third instar in both backgrounds was found to be identical to expression in the normal background. This suggests that the region giving rise to the wing is present, although the wing does not form in *scalloped* and *vestigial* mutants. This is consistent with the *vestigial* enhancer, directed by *apterous*, acting to determine the position within the disc which will eventually attain the identity of wing margin.

The *P* element transposon: Some wild populations of *Drosophila melanogaster* have chromosomes that can generate recombination in the germline of male offspring of crosses between wild males and females from established laboratory stocks (for review see ENGELS 1989). Male recombination and a number of other biological phenomena, including an increased rate of mutation and chromosomal aberrations in the progeny of the F1 males and female gonadal sterility in the F1 females, was termed hybrid dysgenesis (KIDWELL *et al.* 1977). The wild strains and laboratory strains are referred to *P* and *M* strains respectively, to indicate the differences between paternal and maternal activity in dysgenic crosses. Repression of hybrid dysgenesis in *P* strain oocytes is termed cytotype dependency (ENGELS 1979). Once the biological implications of hybrid dysgenesis became established, the molecular basis was soon determined. Mutants of the *white* (*w*) locus were generated during a dysgenic cross (RUBIN *et al.* 1982). The *white* gene had been previously cloned, which allowed evaluation of the nature of these *white* mutants. The dysgenic *white* alleles were demonstrated to possess insertions of different sizes: the *P* elements. With the cloning of the *P* element, hybrid dysgenesis was quickly correlated with the transposition of multiple *P* elements and indicated that the physical manifestation of hybrid dysgenesis is the result of *P* element activity (BINGHAM *et al.* 1982). The availability of cloned *P* element sequences then allowed the cloning of genes (ie. SEARLES *et al.* 1982), including *vestigial* (WILLIAMS and BELL 1988) by transposon tagging and isolation of adjacent genomic sequences through molecular means.

The complete *P* element, 2.9 kb in length, when injected into M strain preblastoderm embryos, could result in strains which expressed the complete *P* cytotype (SPRADLING and RUBIN 1982). The 2.9 kb element can cause transposition of itself and other *P* elements. Sequence analysis established several specific landmarks: 31 basepair terminal inverted repeats, eleven basepair internal repeats and four long open reading frames which encode both the *P* element transposase and the somatic repressor (O'HARE and RUBIN 1983; KARESS and RUBIN 1984; MISRA and RIO 1990). Several hundred base pairs at each end, including the 31 basepair inverted repeats, are necessary for efficient transposition (MULLINS *et al.* 1989). The transposase binds at sites internal to the inverted repeats (KAUFMAN *et al.* 1989) and the inverted-repeat binding protein (IRBP) binds to the termini (RIO and RUBIN 1988). The *P* element is extensively used in Drosophila research as a carrier of genetic material in germline transformation (SPRADLING and RUBIN 1982), as a detector of transcriptional control activity (O'KANE and GEHRING 1987) and as a transposon tag in the molecular characterization of genomic regions (SEARLES *et al.* 1982).

To facilitate the molecular cloning of the *vestigial* genomic region, *P* element transposon tagging was utilized (WILLIAMS and BELL 1988). A cross of Oregon-R (M cytotype) females to males of a *+/SM5* (*P* cytotype) stock, derived from the $\pi2$ line, produced dysgenic F1 +/*SM5* males which were then mated to *vestigial*^{1} flies (M cytotype) and the F2 progeny were scored for wing phenotypes. Of nearly one hundred individual flies bearing wing defects, only one was found to be allelic to *vestigial.* The remaining mutants, mostly alleles of X-linked genes and, perhaps a number of potentially tagged enhancers of *vestigial*, were not analyzed further. The *vestigial* mutant isolated in this screen, designated *vestigial*^{21} , is a cryptic allele which produces a wild type wing phenotype as a homozygote and an intermediate strap to scalloped wing in heterzygous combination with classic *vestigial* alleles (in an M cytotype genetic background). This strain has been extensively used in the study of both *vestigial* and *P* element biology

The cryptic nature of the *vestigial*^{21} allele has readily lent itself to the isolation of extreme derivatives (WILLIAMS *et al.* 1988a). A *P* cytotype version of the *vestigial*^{21} strain was reconstructed by repetitive crossing to a *P* strain balanced for the second chromosome (by the multiply-inverted *SM5* chromosome) and reisolation of the original second chromosome. To produce dysgenic F1 males, *vestigial*^{21} males of the P cytotype were crossed to M cytotype *vestigial*^{21} females. These were then backcrossed to *vestigial*^{21} M cytotype females. Of 15,000 F2 flies screened, 33 possessed a *vestigial* phenotype of which 12 produced progeny. Seven are independent in origin and are confirmed *vestigial* alleles. Two lethal alleles, *vestigial*$^{21-9}$ and *vestigial*$^{21-12}$, were identified as deletions of the *P* element and adjacent genomic regions, and the lethality is not associated with the *vestigial* locus. The remaining five derivatives were viable as homozygotes and fell into three classes: internal deletions (*vestigial*$^{21-7}$ and *vestigial*$^{21-8}$), an adjacent deletion (*vestigial*$^{21-4}$) and large *P* element recruitment events (*vestigial*$^{21-3}$ and *vestigial*$^{21-6}$). Two independent spontaneous revertants were also isolated. The *vestigial*$^{21-4R}$ tertiary derivative exhibited strap wings, compared to the classical *vestigial* wings of *vestigial*$^{21-4}$. In addition to the 36 base pair deletion 3' of the insertion point associated with the *vestigial*$^{21-4}$ allele, the revertant has a deletion of the 5' region (position 16 to 239) of the *P* element. The revertant *vestigial*$^{21-7R}$, a further insert to *vestigial*$^{21-7}$ of a 1.1 kb *P* element, acts as a very weak allele. This recruitment is unusual because the 5' side appears to be an exact replacement of recipient with donor sequences but the 3' end is composed of the complete 3' *P* element terminus inserted beside the 100th basepair of the target *vestigial*$^{21-7}$ *P* element. The abilities of a *P* element to become smaller and larger appears to be the result of an interesting modification of the transposition mechanism.

One *P* element may replace another during hybrid dysgenesis (GEYER *et al.* 1988; WILLIAMS *et al.* 1988a). The *P* element-directed targeting techniques, gene targeting (GLOOR *et al.* 1991; JOHNSON-SCHLITZ and ENGELS 1993; NASSIF *et al.* 1994) and targeted transposition (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994), apply this aspect of *P* element biology to place known sequences at predetermined genomic locations. The initiating event is the production of a double-strand break caused by *P* element excision through the cut-and-paste transposition mechanism (KAUFMAN and RIO 1992). The reversion of a *P* element insertion at *white* to wild type was shown to be more efficient in the presence a homologous chromosome (ENGELS *et al.* 1990). This was followed by the suggestion that a double-strand DNA break generated by *P* element excision is repaired by copying information from a homologous sequence. This has been confirmed by the utilization of altered *white* templates (GLOOR *et al.* 1991; JOHNSON-SCHLITZ and ENGELS 1993; NASSIF *et al.* 1994). The flow of information is unidirectional, as no crossing over or template alteration is observed, which suggests that the donor and recipient are not covalently linked during the repair process. Since donor sites scattered throughout the genome can act as templates in the repair of *white*, a search for sequence homology must be carried out. This may be explained by the synthesis-dependent strand annealing model (NASSIF *et al.* 1994). According to this model, DNA synthesis is initiated at the free ends of a double-strand gap to copy a template (possibly ectopic) and is completed by annealing of the strands followed by additional DNA synthesis to finish the repair of the gap.

P element transposition is repressed by products encoded by the *P* elements themselves: the *P* cytotype (ENGELS 1989; RIO 1990; RIO 1991). The transposase has been shown to repress transcription *in vitro* by binding to a site at the 5' end of the *P* element (KAUFMAN *et al.* 1989, KAUFMAN and RIO 1991) to prevent the binding of the transcription factor (TFIID) at an overlapping site. A 66 kD somatic repressor, a truncated version of the transposase protein, has been isolated and apparently mimics the binding activity of the transposase (MISRA and RIO 1990). Modified *P* elements that encode the 66 kD repressor protein did not initially display maternal repression of *P* element activity as monitored by the *singed*^{weak} test (MISRA and RIO 1990). However, mobilization to several new positions did result in significant maternal repression (MISRA *et al.* 1993). The repression of *P* element transposition by the *P* cytotype appears to be achieved via the control of *P* element transcription in the germline.

Control of gene activity in the germline aside, the products of the *P* element may influence gene expression of a number of genes including *vestigial* (WILLIAMS *et al.* 1988b). Genes, other than the repressor/transposase transcription unit, that have been placed under the transcriptional control of *P* element sequences are presumably susceptible to, at the very least, a subset of the mechanisms that direct *P* element transcription. Several *singed* alleles harbouring defective *P* element inserts are suppressed in the *P* cytotype (ROBERTSON and ENGELS 1989). The *P* cytotype can also influence the expression of the *white* gene inserted within a *P* element but is strongly position dependent (COEN 1990). The activity of a *P-lacZ* reporter gene has been shown to be controlled by the *P* element regulatory products such that *lacZ* expression is greatly reduced in a *P* background (LEMAITRE and COEN 1991). *P-lacZ* expression that reflects cytotype was observed in recessive female-sterile enhancer trap lines that have ß-galactosidase activity in germline tissues (LEMAITRE *et al.* 1993). In this study, crosses to *P* strains revealed high levels of ß-galactosidase activity in the germline of the offspring of *P* males and *M* females and repression in the progeny of the reciprocal cross. Clearly, non-*P* element genes may be placed under the influence of *P* element proteins.

The phenotypes of several of the *P* alleles of *vestigial* were evaluated to determine the ability of the *P* cytotype to suppress the associated *vestigial* wing defects observed in the M cytotype (WILLIAMS et al 1988b). Many of the *P* element alleles tested, *vestigial*$^{21-4}$, *vestigial*$^{21-4R}$, *vestigial*$^{21-7}$, *vestigial*$^{21-8}$ and *vestigial*$^{21-9}$, as well as *vestigial*^{1}, are not suppressed in the *P* cytotype. The original *P* element allele, *vestigial*^{21}, in heterozygous combination with *vestigial*^{1}, is completely suppressed such that the intermediate wing phenotype seen in the M cytotype is transformed to wild type in the *P* cytotype. Similarly, the scalloped wing phenotype of *vestigial*$^{21-7R}$ observed in the *M* cytotype is completely suppressed in the *P* cytotype. The *vestigial*$^{21-3}$ and *vestigial*$^{21-6}$ alleles are also influenced by cytotype and display the classic *vestigial* phenotype in an M cytotype and complete wings in a *P* cytotype background. Unlike the *P* cytotype's influence upon hybrid dysgenesis, the suppression of the *vestigial* *P* element alleles is not maternally inherited, and is observed when *P* elements are inherited from either parent. The sensitivity of the *vestigial*$^{21-3}$ allele to the presence of the *P* cytotype has been exploited to identify *P* element type I repressors (GLOOR et al 1993). It should be noted that cytotype, by definition, is restricted to the germline and suppression of the *vestigial* phenotype in the soma is a related phenomenon which apparently depends upon some of the same molecular components. Perhaps a term such as *P* element repression would better describe the somatic effects discussed above.

The studies reported here examine and exploit properties of the *vestigial* gene, the *P* element transposon and their interactions in the *vestigial*^{21} allele and its derivatives. Previous studies have demonstrated that the *P* element of *vestigial*^{21} can be altered and that such changes can influence the phenotypes of the derivatives (WILLIAMS *et al.*, 1988a;b). The current investigations were initiated to address questions which arose, in part, from the above experiments. The ability of *vestigial*^{21} to recruit other *P* elements to its genomic position (WILLIAMS *et al.*, 1988a) allowed the targeting of an "enhancer-trap" *P* element, *P[lacZ; ry*$^{+}$*]*, to the *vestigial* gene. An enhancer trap is a transposon which has been modified *in vitro* to bear a reporter gene under the control of a weak promoter and then returned to the germline (O'KANE and GEHRING 1987). The enhancer trap can be mobilized to new locations such that the reporter gene comes under the control of transcriptional control regions present near its insertion site. The isolation of this unique *vestigial* allele was dependent upon the generation of an extreme *vestigial* phenotype due to the recruitment of the exogenous sequences to the site of the *vestigial*^{21} *P* element. The targeted transposition of a reporter gene construct to a predetermined genomic site is a novel technique which combines targeted transposition (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994) and the use of enhancer traps (O'KANE and GEHRING 1987). The tendency of *P* elements to become internally-deleted such that one or both breakpoints exist within the terminal inverted repeats (O'HARE and RUBIN 1983; SEARLES *et al.* 1986) was further demonstrated through the generation of a large number of deletion derivatives of *vestigial*^{21}. The deletion breakpoints within the termini were found to be located at a position approximately 16 base pairs from the end of the *P* element. The inverted-repeat binding protein (IRBP) is known to interact with this sequence *in vitro* and these results suggest that this protein may have a role in the process that leads to internal deletion. Correlation of the *P* element sequences of *vestigial*^{21} and its derivatives to their respective mutant phenotypes has revealed that *P* element activity may influence *vestigial* expression (WILLIAMS *et al.* 1988a; 1988b). Transcriptional interference at the *vestigial* promoter by the transcription initiated at the *P* promoter may account for the mutant phenotypes, which is a distinct alternative to the models previously proposed to account for the observed phenotypes (WILLIAMS *et al.* 1988a; RIO 1990). The mechanism by which *P* element sequences inhibit *vestigial* transcription may also be responsible for the phenotypes of *P* element alleles of other genes such as *singed* (ROBERTSON and ENGELS 1989). Previous examination of *vestigial* transcription products revealed a 3.8 kb transcript (WILLIAMS *et al.*; 1990a). A number of other transcripts were also demonstrated to hybridize to *vestigial* probes but were attributed to non-specific interactions. Re-examination of the products of *vestigial* transcription indicates that the *vestigial* pre-mRNA undergoes alternative splicing which is contrary to the conclusions drawn from a previous study of *vestigial* transcription (WILLIAMS *et al.* 1990a). This suggests that the expression of *vestigial* may be dependent upon differential processing of the message. The investigations herein contribute to the understanding of *P* element biology, present a mechanism by which transposons may influence a host gene and further the study of a gene that is important in Drosophila development.

LITERATURE CITED

ALEXANDROV, I. D., and M. V. ALEXANDROVA, 1987 A new *nw* allele and interallelic complementation at the *vg* locus of *Drosophila melanogaster*. Drosophila Inf. Serv. **66**: 11.

BELL, L. R., E. M. MAINE, P. SCHEDL and W. T. CLINE, 1988 *Sex-lethal* , a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell **55**: 1037-1046.

BINGHAM, P. M., M. G. KIDWELL and G. M. RUBIN, 1982 The molecular basis of *P-M* hybrid dysgenesis: the role of the *P* element, a *P*-strain-specific transposon family. Cell **29**: 995-1004

BOWNES, M., and S. ROBERTS, 1981 Regulative properties of wing discs from the *vestigial* mutant of *Drosophila melanogaster*. Differentiation **18**: 89-96.

BRIDGES, C. B., and T. H. MORGAN, 1919 The second chromosome group of mutant characters. Carnegie Inst. Washington Publ. **278**: 123-304.

CAMPBELL, S., M. INAMDAR, V. RODRIGUES, V. RAGHAVAN, M. PALAZZOLO and A. CHOVNICK, 1992 The *scalloped* gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev. **6**: 367-379.

COEN, D., 1990 *P* element regulatory products enhance $zeste^{1}$ repression of a $P[white^{duplicated}]$ transgene in *Drosophila melanogaster*. Genetics **126**: 949-960.

COHEN, B., M. E. MCGRIFFIN, C. PFEIFLE, D. SEGAL and S. M. COHEN, 1992 *apterous*, a gene required for imaginal disc development in Drosophila encodes a member of the LIM family of developmental regulatory proteins. Genes Dev. **6**: 715-729.

COHEN, S. M., 1993. *Imaginal disc development*, pp. 747-841 in *Drosophila Development*, edited by A. MARTINEZ-ARIAS and M. BATE. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

COHEN, S. M., and G. JURGENS, 1989 Proximal-distal pattern formation in Drosophila: cell autonomous requirement for *Distal-less* gene activity in limb devlopment. EMBO J. **8**: 2045-2055.

ENGELS, W. R., 1979 Extrachromosomal control of mutability in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA. **76**: 4011-4015.

ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 437-484 in *Mobile DNA.*, edited by D.E. BERG and M.M. HOWE. American Society for Microbiology, Washington DC.

ENGELS, W. R., D. M. JOHNSON-SCHLITZ, W. B. EGGLESTON and J. SVED, 1990 High-frequency *P* element loss in Drosophila is homolog dependent. Cell **62**: 515-525.

FRISTROM, D., 1968 Cellular degeneration in wing development of the mutant *vestigial* of *Drosophila melanogaster*. J. Cell Biol. **39**: 488-491.

FRISTROM, D., 1969 Cellular degeneration in the production of some mutant phenotypes in *Drosophila melanogaster*. Mol. Gen. Genet **103**: 363-379.

GEYER, P. K., K. L. RICHARDSON, V. G. CORCE and M. M. GREEN, 1988. Genetic instability in *Drosophila melanogaster*: *P*-element mutagenesis by gene conversion. Proc. Natl. Acad. Sci. USA **85**: 6455-6459.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. R. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in Drosophila via *P* element-induced gap repair. Science **253**: 1110-1117.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS, W. K. BENZ, H. M. ROBERTSON and W. R. ENGELS, 1993 Type I repressors of *P* element mobility. Genetics **135**: 81-95.

GREEN, M. M., 1946 A study in gene action using different dosages and alleles of *vestigial* in *Drosophila melanogaster*. Genetics **31**: 1-20.

GREEN, M. M., and C. P. OLIVER, 1940 The action of certain mutants upon the penetrance of heterozygous *vestigial* wing in *Drosophila melanogaster*. Genetics **25**: 584-592.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131**: 917-927.

HESLIP, T. R., and R. B. HODGETTS, 1994 Targeted transposition at the *vestigial* locus of *Drosophila melanogaster*. Genetics **138**: 1127-1135.

HODGES, D., and S. I. BERNSTEIN, 1994 Genetic and biochemical analysis of alternative RNA splicing. Adv. Genet. **31**: 207-281.

JOHNSON-SCHLITZ, D. M., and W. R. ENGELS, 1993 *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila melanogaster*. Mol. Cell Biol. **13**: 7006-7018.

KARESS, P. D., and G. M. RUBIN, 1984 Analysis of *P* transposable element functions in Drosophila. Cell **38**: 135-146.

KAUFMAN, R. E., R. F. DOLL and D. C. RIO, 1989 Drosophila *P* element transposase recognizes internal *P* element DNA sequences. Cell **59**: 359-371.

KAUFMAN, R. E., and D. C. RIO, 1991 Drosophila *P*-element transposase is a transcriptional repressor *in vitro*. Proc. Natl. Acad. Sci. USA **88**: 2613-2617.

KAUFMAN, R. E., and D. C. RIO, 1992 *P* element transposition *in vitro* proceeds by a cut and paste mechanism and uses GTP as a cofactor. Cell **69**: 27-39.

KIDWELL, M. G., J. F. KIDWELL and J. A. SVED, 1977 Hybrid dysgensis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics **86**: 813-833.

LASKI, F. A., D. C. RIO and G. M. RUBIN, 1986 Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44**: 7-19.

LEMAITRE, B., and D. COEN, 1991. *P* regulatory products repress *in vivo* the *P* promoter activity in *P-lacZ* fusion genes. Proc. Natl. Acad. Sci. USA **88**: 4419-4423.

LEMAITRE, B., S. RONSSERAY and D. COEN, 1993 Maternal repression of the *P* promoter in the germline of *Drosophila melanogaster*: A model for the *P* cytotype. Genetics **135**: 149-160.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

MATTOX, W., and B. S. BAKER, 1991 Autoregulation of the splicing of transcripts from the *transformer-2* gene of Drosophila. Genes Dev. **5**: 786-796.

MCKEOWN, M., 1992 Alternative mRNA splicing. Ann. Rev. Cell Biol. **8**: 133-155.

MISRA, S., and D.C. RIO, 1990 Cytotype control of Drosophila *P* element transposition: the 66kD protein is a repressor of transposase activity. Cell **62**: 269-384.

MISRA, S., R. M. BURATOWSKI, T. OHKAWA and D. C. RIO, 1993 Cytotype control of *Drosophila melanogaster P* element transposition: genomic position determined maternal repression. Genetics **135**: 785-800.

MORGAN, T. H., 1911 The origin of nine wing mutations in Drosophila. Science **33**: 496-499.

MULLINS M. C., D. C. RIO and G. M. RUBIN, 1989 Cis-acting DNA sequence requirements for *P* element transposition. Genes Dev. **3**: 729-738.

NASSIF, N., J. PENNEY, S. PAL, W. R. ENGELS and G. B. GLOOR, 1994 Efficient copying of nonhomologous sequences from ectopic site via *P*-element-induced gap repair. Mol. Cell Biol. **14**: 1613-1625.

O'BROCHTA, D., and P. J. BRYANT, 1983 Cell degeneration and elimination in the imaginal wing disc, caused by the mutations *vestigial* and *Ultravestigial* of *Drosophila melanogaster*. Roux's Arch. Dev. Biol. **192**: 285-294.

O'HARE, K., and G. M. RUBIN, 1983 Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**: 25-35.

O'HARE, K., A. DRIVER, S. MCGRATH and D. M. JOHNSON-SCHLITZ, 1992 Distribution and structure of cloned *P* elements from the *Drosophila melanogaster P* strain π2. Genet. Res., Camb. **60**: 33-41.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection i*n situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84**: 9123-9127.

RIO., D. C., 1990 Molecular mechanisms regulating Drosophila *P* element transposition. Ann. Rev. Genet. **24**: 543-578.

RIO, D. C., 1991 Regulation of Drosophila *P* element transposition. Trend. Genet. **7**: 282-287.

RIO, D. C., and G. M. RUBIN, 1988 Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element. Proc. Natl. Acad. Sci. USA **85**: 8929-8933.

ROBERTSON, H. M., and W. R. ENGELS, 1989 Modified *P* elements that mimic the *P* cytotype in *Drosophila melanogaster*. Genetics **123**: 815-824.

RUBIN, G. M., M. G. KIDWELL and P. M. BINGHAM, 1982 The molecular basis of *P-M* hybrid dysgenesis: the nature of induced mutations. Cell **29**: 987-994.

SEARLES L. L., R. S. JOKERST, P. M. BINGHAM, R. A. VOELKER and A. L.GREENLEAF, 1982 Molecular cloning of sequences from a Drosophila *RNA polymerase II* locus by *P* element transposon tagging. Cell **31**: 585-592.

SEARLES L. L., A. L.GREENLEAF, W. E. KEMP and R. A. VOELKER, 1986 Sites of *P* element insertion and structures of *P* element deletions in the 5' region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6**: 3312-3319.

SPRADLING, A. C., and G. M. RUBIN, 1982 Transposition of cloned *P* elements into Drosophila germ line chromosomes. Science **218**: 341-347.

WILLIAMS, J. A., 1989 The *vestigial* locus of *Drosophila melanogaster*. Ph.D. thesis, University of Alberta.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**: 1355-1363.

WILLIAMS, J. A., and S. B. CAROLL, 1993 The origin, patterning and evolution of insect appendages. Bioessays **15**: 567-578.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988a Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988b Suppressible *P*-element alleles of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **212**: 370-374.

WILLIAMS, J. A., A. L. AKTIN and J. B. BELL, 1990a The functional organization of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **221**: 8-16.

WILLIAMS, J. A., I. M. SCOTT, A. L. AKTIN, B. J. BROOK, M. A. RUSSELL and J. B. BELL, 1990b Genetic and molecular analysis of vg^U and vg^W: Two dominant *vg* alleles associated with gene fusions in Drosophila. Genetics **125**: 833-844.

WILLIAMS, J. A., J. B. BELL and S. B. CARROLL, 1991 Control of Drosophila wing and haltere development by the nuclear *vestigial* gene product. Genes Dev. **5**: 2481-2495.

WILLIAMS, J. A., S. W. PADDOCK and S. B. CARROLL, 1993 Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. Development **117**: 571-584.

WILLIAMS, J. A., S. W. PADDOCK, K. VORWERK and S. B. CARROLL, 1994 Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature **368**: 299-305.

Chapter 2

Targeting of an enhancer trap to *vestigial*

A version of this chapter has been published. STAVELEY *et al*. 1994. *Developmental Biology*. 165: 290-293.

The *vestigial* gene product is required for the development of the adult wing. The wing phenotypes of *vestigial* mutants range from a small classic *vestigial* wing, through strapped, scalloped and nicked to wild-type wing (WILLIAMS and BELL 1988; LINDSLEY and ZIMM 1992). Pleiotropic effects associated with some alleles of *vestigial* include haltere reduction, erect postscutellar bristles, lowered viability and female sterility. The gene spans at least 16 kb within polytene chromosome section 49D and produces a single 3.8 kb transcript (WILLIAMS *et al.* 1990). The expression pattern of *vestigial* indicates that it plays an important role in the global development as a 'pro-wing' gene in the wing imaginal disc (WILLIAMS *et al.* 1993). The protein has 453 amino acids and may be involved in protein-protein interactions, but the mechanism by which the *vestigial* protein acts is unknown (WILLIAMS *et al.* 1991; 1994). The low rate of recovery of *P* element alleles of *vestigial* (1 in 2 X 10^6; WILLIAMS and BELL 1988) has encouraged the approach described below to isolate an enhancer trap at *vestigial*.

Several enhancer trap constructs have been developed and a large number of enhancer trap lines have been generated in the study of developmentally important genes (ie. O'KANE and GEHRING 1987; BELLEN *et al.* 1989). However, the difficulty of generating enhancer trap alleles of genes in which *P* element alleles are not readily isolated has not been adequately addressed. Some *P* elements have the ability to attract another *P* element to their site of insertion (GEYER *et al.* 1988; WILLIAMS *et al.* 1988). Our work is based on the *P* element-induced *vestigial*^{21} allele, at which targeted transposition of *P[Ddc]* has been shown to occur (HESLIP *et al.* 1992). Targeted gene replacement at the *white* locus seems to occur via excision of the resident *P* element followed by double-strand gap repair duplicating an ectopic copy of *white* (GLOOR *et al.* 1991). If a gap repair process is responsible for *P* element recruitment to *vestigial*^{21}, it must differ from the targeted gene replacements at *white* since the only homology between the target and the ectopic template resides in the *P* element itself. We emphasize this difference by describing *P* element replacements as targeted transpositions and not targeted gene replacements. The objective of this work was to target an "enhancer trap" to *vestigial* and characterize its reporter gene activity. The generation of an allele of this nature could provide an alternative approach by which aspects of the transcriptional control of *vestigial* may be examined.

MATERIALS AND METHODS

Drosophila Strains: The *P[lacZ;ry]*^{c49}; *b cn vg*^{21}; *ry* line possesses an internally deleted *P* element inserted in the 5' untranscribed region of *vestigial* (WILLIAMS and BELL 1988) and an X-linked enhancer trap *P* element (O'KANE and GEHRING 1987). The *w*; *Sb P[Δ2-3,ry](99B) e/TM6, e* stock provided the stable source of transposase (ROBERTSON *et al.* 1988). *Df(2R)vg*^{B}*/SM5* is described in LINDSLEY and ZIMM (1992). All flies were maintained at 22°C on standard yeast/glucose medium (NASH and BELL 1968).

Targeting Scheme: *P[lacZ;ry]*^{c49}; *b cn vg*^{21}; *ry* females were crossed to *w*; *Sb P[Δ2-3,ry] e/TM6, e* males to produce dysgenic males. Dysgenic males of the genotype *P[lacZ;ry]*^{c49}; *b cn vg*^{21}*/+*; *Sb P[Δ2-3,ry](99B) e/ry* were mated *en masse* to 20 - 25 *b cn vg*^{21} females. Individuals heterozygous for *vg*^{21} and a strong *vestigial* allele have nicked or scalloped wings but *vg*^{21} homozygotes display a wild type phenotype. Selection was based upon the assumption that recruitment of *P[lacZ;ry*$^{+}$*]* sequences to *vg*^{21} would result in a non-cryptic allele of *vestigial* (HESLIP *et al.* 1992). Offspring displaying any wing-margin defects were crossed to *Df(2R)vg*^{B}*/SM5* individuals, the hemizygous offspring were inspected for wing phenotypes and lines were established. The allele designation reflects the number of the *en masse* mating vessel such that the lines labeled *vg*$^{21\text{-}n.1}$ and *vg*$^{21\text{-}n.2}$ (where n=1 to 200) represent different isolates for a single *en masse* mating and therefore may be the result of a single premeiotic dysgenic event.

DNA manipulations: Genomic DNA was isolated from derivatives of *vg*^{21}, digested with restriction enzymes and fractionated by electrophoresis in 0.8% agarose/TBE gels according to standard methodologies (SAMBROOK *et al.* 1992). The DNA was transferred by Southern blotting to GeneScreen Plus nylon membranes (Du Pont) in accordance with the manufacturer's instructions and hybridization was carried out at 42°C in 50% formamide solution. The *vestigial* locus probes were derived from cloned sequences (WILLIAMS and BELL 1988). Restriction fragments were isolated by agarose gel electrophoresis, purified by GeneCleanII (Bio 101 Inc.) and oligo-labeled in the presence of ^{32}P dCTP (ICN) to generate radioactive probes for hybridization. The blots were washed under stringent conditions and exposed to X-ray film. A 0.7 kb *Eco*RI fragment located at the 5' end of the gene is increased in size to 1.4 kb in the *vg*^{21} mutant. The *lacZ* probe was generated from a 3 kb *Bam*HI fragment of pMGb, a derivative of pHZ50PL (HIROMI and GEHRING 1987). The PCR fragments were generated from the pairwise combinations of *vestigial* and *P* element primers (see FIGURE 2.1c).

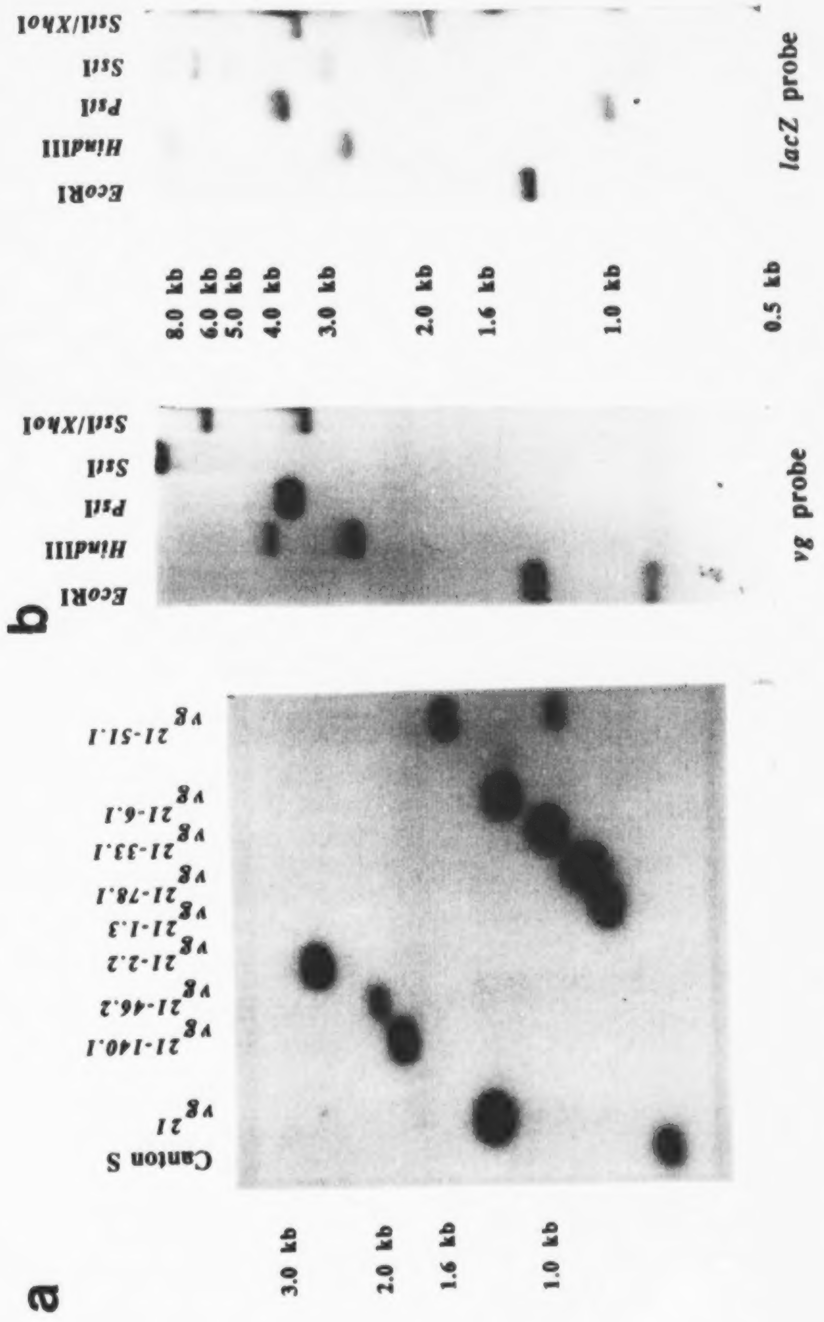
Histochemical Staining: Third instar larvae were dissected in phosphate buffered saline (PBS), fixed in PBS + 0.25% glutaraldehyde for 30 min and rinsed with PBS. The dissected larval material was then stained overnight at 37°C in an X-gal solution comprised of 10 mM NaH_2PO_4/Na_2HPO_4(pH 7.2) 150mM NaCl 1mM $MgCl_2{\cdot}6H_2O$, 3.1mM $K_4[Fe^{II}(CN)_6]$, 3.1mM $K_3[Fe^{III}(CN)_6]$ and 0.3% Triton X-100 plus the addition to 1.5 - 2.0% of an X-gal (8% in dimethylformamide) just prior to incubation (BELLEN *et al.* 1989).

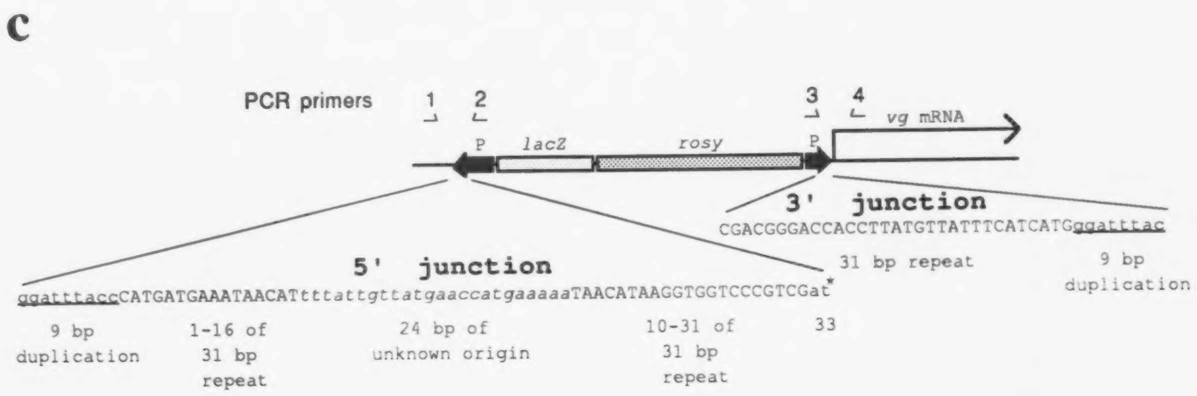
RESULTS AND DISCUSSION

Utilizing the mating scheme outlined above strong alleles of *vestigial* were derived from *vestigial*^{21}. Of 49,616 chromosomes scored, approximately 80 individuals displaying wing margin defects were isolated and 35 lines were established. In the end 15 independently generated lines were subjected to further analysis. All the strong alleles derived from *vestigial*^{21} and analyzed by Southern blot hybridization revealed alterations of the 1.4 kb *Eco*RI fragment when probed with a 5' *vestigial* fragment (FIGURE 2.1a). The *vestigial*^{21} allele contains a 687 bp internally-deleted *P* element inserted into a 0.7 kb *Eco*RI fragment in the 5' region of the gene (WILLIAMS and BELL 1988). The deletions in *vestigial*$^{21\text{-}140.1}$, *vestigial*$^{21\text{-}46.2}$ and *vestigial*$^{21\text{-}2.2}$ remove an adjacent *Eco*RI restriction site (data not shown). In addition, the alleles *vestigial*$^{21\text{-}1.3}$, *vestigial*$^{21\text{-}78.1}$, *vestigial*$^{21\text{-}33.1}$ and *vestigial*$^{21\text{-}6.1}$, harbour deletions within the relevant *Eco*RI fragment. While 14 of the 15 lines have deletions, Southern analysis of *vestigial*$^{21\text{-}51.1}$revealed an *Eco*RI restriction pattern in which the 1.4 kb fragment of *vestigial*^{21} had been altered to give two bands. Further Southern analysis of *vestigial*$^{21\text{-}51.1}$ revealed *vestigial* and *lacZ* sequences to be colinear in this mutant (FIGURE 2.1b).

Sequenced PCR products derived from *vestigial*$^{21\text{-}51.1}$, which include the two junctions between *vestigial* DNA and the targeted enhancer trap (FIGURE 2.1c), show that the enhancer trap has replaced the resident *P* element and its orientation is the same as that of the original *P* element. The 3' terminus is intact but the 5' terminus consists of an interrupted 31 base pair repeat. The existence of a polymorphism at position 33 differentiates the resident *P* element (33=A) from the donor and indicates replacement of resident by donor sequences at this point. No individuals displaying both the *rosy* and

FIGURE 2.1. Identification of a recruited enhancer trap at *vestigial* by Southern analysis. (a) A genomic Southern blot of Canton S, vg^{21} and selected derivatives of vg^{21}. Genomic DNA was digested with *Eco*RI, fractionated on a 1% agarose gel, transferred to a GeneScreen Plus membrane, hybridized to a 700 bp *Eco*RI vg^{+} fragment spanning the insertion site of the *P* element of vg^{21} and subjected to autoradiography. The deletions in $vg^{21\text{-}140.1}$, $vg^{21\text{-}46.2}$ and $vg^{21\text{-}2.2}$ remove an adjacent *Eco*RI restriction site, while the alleles $vg^{21\text{-}1.3}$, $vg^{21\text{-}78.1}$, $vg^{21\text{-}33.1}$, and $vg^{21\text{-}6.1}$ harbour deletions within the 1.4 kb *Eco*RI fragment. $vg^{21\text{-}51.1}$ revealed an *Eco*RI restriction pattern in which the 1.4 kb fragment of vg^{21} has been altered to give two bands of 1.5 and 0.9 kb. (b) A genomic Southern blot of $vg^{21\text{-}51.1}$ (renamed vg^{lacZ1}) hybridized with *vestigial* and *lacZ* sequences. Genomic DNA was digested with *Eco*RI, *Hin*dIII, *Pst*I, *Sst*I and *Sst*I/*Xho*I and probed as above with the 700 bp *Eco*RI vg^{+} fragment. The membrane was stripped of the vg^{+} fragment and rehybridized with a *lacZ* probe. Co-migrating bands which hybridized to both probes, representing the junction fragments, include a 1.5 kb *Eco*RI fragment (lane 1), a 2.9 kb *Hin*dIII fragment (lane 2), a 4.1 kb *Pst*I fragment (lane 3) and a 3.8 kb *Sst*I/*Xho*I fragment (lane 5). (c) A diagram of the targeted enhancer trap inserted at the 5' end of the *vestigial* gene. The DNA sequence of the PCR fragments spanning the 5' and 3' junctions was determined. The products produced by primer pair 1 and 2 and primer pair 3 and 4 indicated the orientation of the insert. Nucleotides found in the 31 bp inverted repeat of the *P* element are capitalized. The 3' junction remains identical to that of the original allele and the terminal repeat is intact. Examination of the 5' terminus of the *P* element reveals a complex sequence consisting of the first sixteen nucleotides of the 5' end of the P element, 24 bases of undetermined origin (italicized) and *P* element sequences starting at nucleotide 10 and including the donor polymorphism at nucleotide 33(*). Both 9 bp duplication sequences present in vg^{21} were maintained (underlined).


a
3.0 kb
2.0 kb
1.6 kb
1.0 kb
Canton S
vg^{21}
$vg^{21\text{-}140.1}$
$vg^{21\text{-}46.2}$
$vg^{21\text{-}2.2}$
$vg^{21\text{-}1.3}$
$vg^{21\text{-}78.1}$
$vg^{21\text{-}33.1}$
$vg^{21\text{-}6.1}$
$vg^{21\text{-}51.1}$
b
EcoRI
HindIII
PstI
SstI
SstI/XhoI
vg probe
8.0 kb
6.0 kb
5.0 kb
4.0 kb
3.0 kb
2.0 kb
1.6 kb
1.0 kb
0.5 kb
EcoRI
HindIII
PstI
SstI
SstI/XhoI
lacZ probe


C
PCR primers
1
2
3
4
P
lacZ
rosy
P
vg mRNA
3' junction
CGACGGGACCACCTTATGTTATTTCATCATGggatttac
31 bp repeat
9 bp duplication
5' junction
ggatttaccCATGATGAAATAACATtttattgttatgaaccatgaaaaaTAACATAAGGTGGTCCCGTCGat*
9 bp duplication
1-16 of 31 bp repeat
24 bp of unknown origin
10-31 of 31 bp repeat
33

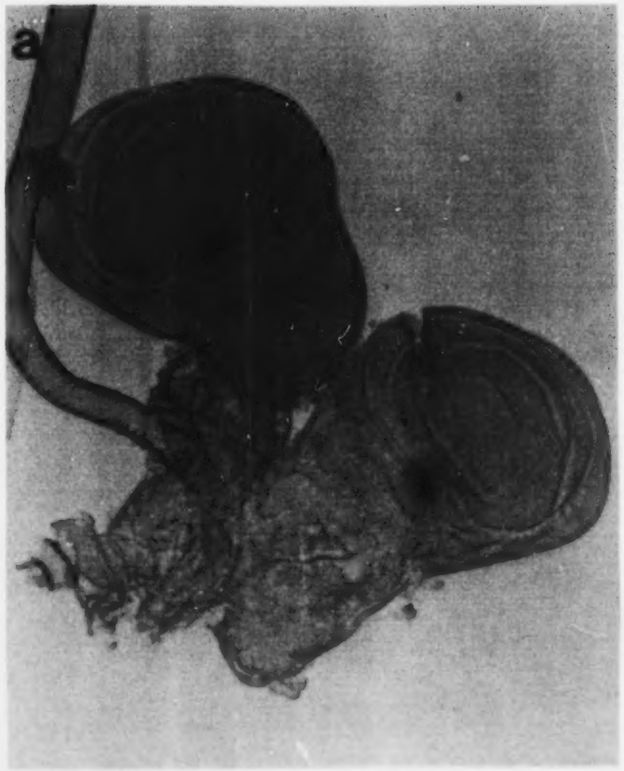
vestigial phenotypes were recovered amoung the F_2 produced from crosses of *vestigial*$^{21-51.1}$ and *rosy* homozygotes. This indicates that the *rosy* function remains intact but we can not rule out that minor alterations of the enhancer trap may have occurred. We conclude that *vestigial*$^{21-51.1}$ represents the replacement of the original *P* element sequences by *P[lacZ;ry*$^+$*]* sequences and have renamed this allele: *vestigial*^{lacZ1}.

The wing and haltere imaginal discs of larvae bearing the targeted enhancer trap were examined for ß-galactosidase activity. Due to the possibility that extensive cell-death in *vestigial* wing discs could obscure expression of the transgene, *vestigial*^{lacZ1}/*vestigial*$^+$ heterozygotes were analyzed. The activity of ß-galactosidase in the wing disc (FIGURE 2.2a) is localized to a central region corresponding to an area of the presumptive dorsal wing surface (fate map: see BRYANT 1975) and represents the typical staining pattern observed in dozens of such discs. While staining of the haltere disc is also localized to a central region homologous to the wing disc site, imaginal leg discs do not show any activity (data not shown). Wing discs from *vestigial*^{lacZ1} homozygotes, which possess a reduced wing pouch and are smaller than the above, also have the same ß-galactosidase activity observed in heterozygotes. Although the reporter gene expression is localized to the primordia of the wing and haltere, the pattern of ß-galactosidase in *vestigial*^{lacZ1} does not resemble the distribution of *vestigial* mRNA or protein in the presumptive wing margin and wing hinge (WILLIAMS *et al.* 1991; FIGURE 2.2b).

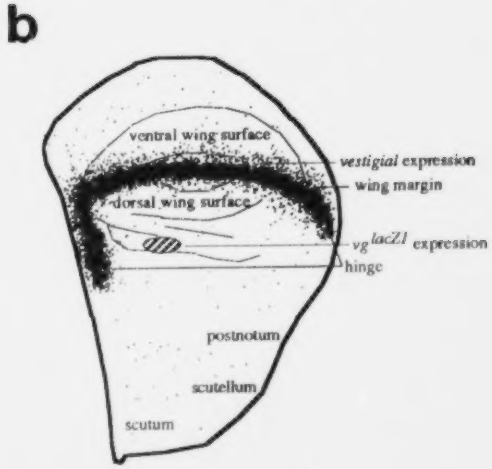
The *vestigial* gene is crucial to the development of the wing and haltere and may play a central role after *wingless* and in cooperation with *scalloped* in the establishment of wing structures (WILLIAMS *et al.* 1993). In the imaginal wing discs of third instar larvae, the nuclear-localized *vestigial* protein is expressed to its maximum level along the length of the presumptive wing margin, decreasing in sharp gradients on both the dorsal and ventral side of the margin (WILLIAMS *et al.* 1991; WILLIAMS *et al.* 1993). A deletion of most of the second intron of *vestigial* (*vestigial*^{83b27}) results in vast reduction of wing and haltere expression which indicates that wing and haltere disc specific enhancers are present within this region.

Recently, it has been demonstrated that a 750 basepair fragment of the second intron of *vestigial* can direct ß-galactosidase expression in a complex pattern throughout the wing discs of third instar larvae (WILLIAMS *et al.* 1994). The staining occurs along the presumptive wing margin and hinge region, as does *vestigial* protein expression, but

FIGURE 2.2. ß-galactosidase staining of *vg*^{lacZ1} imaginal discs. Third instar larvae were dissected in phosphate buffered saline, fixed in the presence of glutaraldehyde and incubated with X-gal (MATERIALS AND METHODS). (a) Wing disc expression is localized to a central region corresponding to presumptive dorsal wing surface. (b) A fate map of the imaginal wing disc showing ß-galactosidase staining of *vg*^{lacZ1} (this study) and the distribution of the *vestigial* gene product in wild type (adapted from WILLIAMS *et al.*, 1991).



a



b
ventral wing surface
vestigial expression
wing margin
dorsal wing surface
vg lacZ1 expression
hinge
postnotum
scutellum
scutum

also in the region of the presumptive dorsal wing surface identical to that of *vestigial*^{lacZ1}. Therefore, the targeted enhancer trap at *vestigial* apparently identifies only a subset of the transcriptional control elements which act upon *vestigial*. We are currently investigating the biological significance of this staining pattern. To date, we can only conclude that the pattern is reproducible, disc-specific and never found in wild type wing discs.

Even though the expression of ß-galactosidase in *vestigial*^{lacZ1} is not the same as the *vestigial* protein in wild type discs, the reporter construct may be responding to positional information signals important in the development of the wing. In cultured wing disc fragments, the presence of a central region of the presumptive dorsal wing surface allowed the regeneration of the complete disc and its absence induced duplication of pattern elements (BRYANT 1975; BRYANT *et al.* 1981). Expression of ß-galactosidase in *vestigial*^{lacZ1} correlates well with the region of wing that led to regeneration in the above experiment. However, the importance of this correlation, if any, remains to be established.

This analysis demonstrates that the enhancer trap has been recruited to the site of the existing *P* element. This is a useful extension of observations in which natural *P* elements and *P[Ddc]* constructs (HESLIP *et al.* 1992) have been targeted to *vestigial*^{21}. The frequency of the targeted *P[lacZ]* transposition (1 in 49,616) was about 5% of that observed for an X-linked *P[Ddc]* (HESLIP and HODGETTS, unpublished observations). GLOOR *et al.* (1991) report significant effects of the donor position on targeted replacements.

The *vestigial*^{lacZ1} allele was isolated via a genetic screen that selected for derivatives of *vestigial*^{21} which display a weak *vestigial* phenotype in heterozygous combination with *vestigial*^{21}. The recruited sequences are likely involved in the establishment of the strong *vestigial* phenotype of *vestigial*^{lacZ1} homozygotes. The mechanism by which this occurs is unknown.

This is the first report of an enhancer trap targeted to a specific gene. We were able to select for alterations of the resident *P* element by phenotype. However, screens based upon PCR and sib-selection could be devised to detect enhancer trap recruitment at other sites. This would allow the generation of enhancer trap alleles of genes which have demonstrated low rates of *P* element recovery or where screening is impractical. Furthermore, for genes in which multiple *P* element insertions have been recovered at different positions, our targeting methodology might offer an opportunity to recruit enhancer traps to each site.

Note added in preparation of the thesis: Since the publication of a version of this chapter, targeting of an enhancer trap to the *Broad-Complex* locus has been reported (Gonzy-Treboul *et al.*, 1995).

LITERATURE CITED

BELLEN, H. J., C. J. O'KANE, C. WILSON, U. GROSSNIKLAUS, R. K. PEARSON and W. J. GEHRING, 1989 *P*-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev. **3**: 1288-1300.

BRYANT, P. J., 1975 Pattern formation in the imaginal wing disc of *Drosophila melanogaster*: Fate map, regeneration and duplication. J. Exp. Zool.: **193**: 47-78.

BRYANT, S. V., V. FRENCH and P. J. BRYANT, 1981 Distal regeneration and symmetry. Science **212**: 993-1002.

GEYER, P. K., K. L. RICHARDSON, V. G. CORCE and M. M. GREEN, 1988 Genetic instability in *Drosophila melanogaster*: *P*-element mutagenesis by gene conversion. Proc. Natl. Acad. Sci. USA **85**: 6455-6459.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. R. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in Drosophila via *P* element-induced gap repair. Science **253**: 1110-1117.

GONZY-TREBOUL, G., J.-A. LEPESANT and J. DEUTSCH, 1995. Enhancer-trap targeting at the *Broad-Complex* locus of *Drosophila melanogaster*: Genes Dev. **9**: 1137-1148.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131**: 917-927.

HIROMI, Y., and W. J. GEHRING, 1987 Regulation and function of the Drosophila segmentation gene *fushi tarazu*. Cell **50**: 963-974.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

NASH, D., and J. B. BELL, 1968 Larval age and the pattern of DNA synthesis in polytene chromosomes. Can. J. Genet. Cytol **10**: 82-92.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84**: 9123-9127.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ and W. R. ENGELS, 1988 A stable genetic source of *P*-element transposase in *Drosophila melanogaster*. Genetics **118**:461-470.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed.2. Cold Spring Harbor Press, Cold Spring Harbor, New York.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**: 1355-1363.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988 Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

WILLIAMS, J. A., A. L. AKTIN, and J. B. BELL, 1990 The functional organization of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **221**: 8-16.

WILLIAMS, J. A., J. B. BELL and S. B. CARROLL, 1991 Control of Drosophila wing and haltere development by the nuclear *vestigial* gene product. Genes Dev. **5**: 2481-2495.

WILLIAMS, J. A., S. W. PADDOCK and S. B. CARROLL, 1993 Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. Development **117**: 571-584.

WILLIAMS, J. A., S. W. PADDOCK, K. VORWERK and S. B. CARROLL, 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature **368**: 299-305.

Chapter 3

Protected *P* element termini in transposase-induced gap repair

A version of this chapter has been published. STAVELEY *et al.* 1995. Genetics. 139: 1321-1329.

The *P* element of *Drosophila melanogaster* is the transposable element responsible for the phenomenon of *P-M* hybrid dysgenesis (reviewed in ENGELS 1989; RIO 1990). The complete *P* element is 2907 base pairs in length and has terminal 31 base pair inverted-repeats (O'HARE and RUBIN 1983). The element encodes both a transposase (KARESS and RUBIN 1984) and a somatic repressor (MISRA and RIO 1990). Transposition usually generates an eight base pair duplication at the target site (see ENGELS 1989; KAUFMAN and RIO 1992) and it requires the presence of a functional *P* element transposase gene (KARASS and RUBIN 1984) and *P* element ends (MULLINS *et al.* 1989). The process is one of cut-and-paste (KAUFMAN and RIO 1992) and the donor site can remain unaltered, revert to wild type, become internally deleted (ENGELS *et al.* 1990) or recruit sequences from an ectopic site (GLOOR *et al.* 1991). Of interest to us is the relationship between the *P* element activities that lead to internal deletion (ENGELS *et al.* 1990; TAKASU-ISHIKAWA *et al.* 1992) and those specialized activities which include targeted gene replacement (GLOOR *et al.* 1991) and targeted transposition (HESLIP and HODGETTS 1994).

Internally-deleted versions make up the majority of *P* elements within the genomes of *P* strains (O'HARE and RUBIN 1983; O'HARE *et al.* 1992). A model of *P* element transposition, which accounts for internally-deleted elements, has been proposed which invokes double-strand gap repair at the break after excision (ENGELS *et al.* 1990) and synthesis-dependent strand annealing (SDSA) to allow efficient copying of sequences to the gap (NASSIF *et al.* 1994). Complete excision of a *P* element leaves behind a double-strand break which is repaired using the homolog, the sister chromatid, or an ectopic copy as a template (GLOOR *et al.* 1991). According to this model (NASSIF *et al.* 1994), each side of the double-strand break independently invades a template and initiates DNA synthesis from this template. The newly synthesized single strands then anneal to each other and further DNA synthesis completes the repair. This model accounts for a number of the incomplete *P* elements recovered (see O'HARE *et al.* 1992). It is further supported by the analysis of two recombinant *P* elements, one bearing a *white-apricot* construct (containing the *gypsy* retrotransposon) and another containing two direct copies of the flip recombinase target sequence (FRT) flanking a white minigene (KURKULOS *et al.* 1994). In both cases, intervening material between the 276-base pair terminal repeats of *gypsy* and between the 599-basepair FRT sites was preferentially deleted.

The appearance of "filler" sequences at the deletion breakpoints of internally-deleted *P* elements is common in naturally-occurring *P* elements (O'HARE and RUBIN

1983; SEARLES *et al.* 1986; WILLIAMS *et al.* 1988; O'HARE *et al.* 1992) and in internally-deleted derivatives of recombinant *P* elements (TAKASU-ISHIKAWA *et al.* 1992). DNA replication and template slippage was suggested as a possible source of some filler sequences (O'HARE and RUBIN 1983). A different model, based upon a hairpin repair mechanism has also been proposed (TAKASU-ISHIKAWA *et al.* 1992). In their model, the free ends generated by the excision of the *P* element are ligated to form two hairpin structures. Nicking and linearization of the hairpins followed by ligation and DNA synthesis would then introduce extra sequences into the deletion breakpoint. We have found that filler sequences at the breakpoints of internally-deleted *P* elements often resemble short stretches of the *P* element terminus, sometimes in tandem arrays. We favour a replication-slippage model to explain their creation during gap repair.

Herein we describe a set of internally-deleted *P* elements generated, in the presence of a stable source of transposase, from an extant *P* element insertion allele. The *vestigial*^{21} allele (WILLIAMS and BELL 1988) has a wild type phenotype as a homozygote but a strap wing phenotype in combination with most strong alleles of *vestigial*. This as a cryptic allele. In phenotypic selection experiments, we have isolated both strong alleles of *vestigial* and revertants of the haplo-insufficient *vestigial*^{21} hemizygous phenotype. All of the internal deletions generated have a breakpoint within either one or both of the terminal 31 base pair inverted-repeats and most retain the last 16±1 base pairs of the end. This terminal sequence has been demonstrated to bind the inverted-repeat-binding protein (IRBP; RIO and RUBIN 1988). The above model of double-strand gap repair (ENGELS *et al.* 1990; NASSIF *et al.* 1994) does not adequately account for our recovery of internally-deleted *P* elements with breakpoints within the terminal inverted-repeats. We propose that the IRBP acts to maintain the integrity of the *P* element terminus which can then serve as a template for DNA synthesis initiated at the other side of the double-strand gap and, therefore, repair the break.

MATERIALS AND METHODS

Drosophila strains: The $P[lacZ;ry^{+}]^{c49}$; $b\ cn\ vg^{21}$; *ry* line was constructed from $b\ cn\ vg^{21}$, which possesses an internally-deleted *P* element inserted in the 5' untranscribed region of *vestigial* (WILLIAMS and BELL 1988) and $P[lacZ;ry]^{c49}$, an X-linked enhancer trap *P* element (O'KANE and GEHRING 1987). The $P[Ddc^{+}](35C)\ Ddc^{ts2}\ pr\ vg^{21}$ line was produced as previously described (HESLIP and HODGETTS 1994). The *w*; $Sb\ P[\Delta 2\text{-}3,ry^{+}]\ e/TM6,\ e$ stock provided the stable source of transposase (ROBERTSON *et al.*

1988). The other Drosophila strains are described in LINDSLEY and ZIMM (1992). All flies were maintained at 22°C on a standard yeast/glucose medium.

Selection for *vestigial* phenotype: The extreme *vestigial* alleles were generated during experiments to study targeted transposition (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994; STAVELEY *et al.* 1994). $P[lacZ;ry^{+}]^{c49}$; *b cn* vg^{21}; *ry* females were crossed to *w*; $Sb\ P[\Delta 2\text{-}3,ry^{+}]\ e/TM6,\ e$ males to produce dysgenic males of the genotype $P[lacZ;ry^{+}]^{c49}$; *b cn* $vg^{21}/+$; $Sb\ P[\Delta 2\text{-}3,ry^{+}]\ e/ry$. About ten dysgenic males were mated *en masse* to 20 - 25 *b cn* vg^{21} females in pint bottles. Individuals heterozygous for *vestigial*^{21} and a strong *vestigial* allele have nicked or scalloped wings but *vestigial*^{21} homozygotes display a wild type phenotype. Selection was based upon the assumption that recruitment to or deletion of sequences at *vestigial*^{21} would result in a non-cryptic allele of *vestigial* (HESLIP *et al.* 1992). Offspring displaying any wing-margin defects were crossed to $Df(2R)vg^{B}/SM5$ individuals, or *Sp Bl* L^{rm}/CyO and lines were established for further analysis. The lines were inspected for extreme *vestigial* wing phenotypes. The allele designation reflects the number of the *en masse* mating vessel, such that the lines labeled $vg^{21\text{-}n.1}$ and $vg^{21\text{-}n.2}$ (where n=1 to 300) represent different isolates from a single *en masse* mating and therefore may be the result of a single premeiotic dysgenic event. The $vg^{21\text{-}f}$ series were generated in a similar manner in that $P[Ddc^{+}](35C)\ Ddc^{ts2}$ *pr* vg^{21}/CyO; $Sb\ P[\Delta 2\text{-}3,ry^{+}]\ e/+$ dysgenic males were crossed to Ddc^{ts2} *pr* vg^{79d5}/CyO females and offspring displaying a phenotype more severe than vg^{21}/vg^{79d5} were selected (HESLIP and HODGETTS 1994).

Selection for phenotypic revertants of *vestigial*: Revertants of the haplo-insufficient property of *vestigial*^{21} (the scalloped to strapped wing phenotype observed in *vestigial*^{21} hemizygotes) were generated using three different chromosomal constitutions. Dysgenic males bearing the transposase source $P[\Delta 2\text{-}3](99B)$ on chromosome 3 and the *vestigial*^{21}-bearing second chromosome possessed one of three second chromosome homologues; (I) a standard sequence chromosome with a wild type copy of *vestigial*, (II) the multiply inverted balancer chromosome (*CyO*) with a wild type copy of *vestigial* or (III) the deficiency containing chromosome $Df(2R)vg^{B}$, deleted for *vestigial* and several surrounding genes (LASKO and PARDUE 1989). The revertant allele designation reflects the chromosome configuration from which it was selected such that *vestigial*$^{21\text{-}IR1}$ is the first of a series of revertant alleles descended from a dygenic male which in addition to the *vestigial*^{21} chromosome, possessed a standard sequence chromosome.

Selection for altered size of PCR fragment: To verify that the specificity in the positioning of the deletion breakpoints which were isolated in the above experimental procedures (see RESULTS) were not merely the result of an inherent bias of the above methods of selection, a small scale experiment that did not discriminate by phenotype was undertaken. A similar experiment has been carried out at *white* (JOHNSON-SCHLITZ and ENGELS 1993). Dysgenic males of the genetic constitution $vg^{21}/+$; $Sb\ P[\Delta 2\text{-}3,ry^{+}]$ $e/+$ (with and without the presence of the $P[lacZ;ry^{+}]^{c49}$ transposon) were crossed to $Df(2R)vg^{B}/CyO$ females, $Cy\ Sb^{+}$ progeny were crossed to $Df(2R)vg^{B}/CyO$ and lines were established. From these stocks, single flies possessing one of the dysgenic male's second chromosomes in hemizygous combination with the *vestigial* deficiency chromosome were evaluated for PCR fragment size (see **DNA amplification**). Those different from either $vestigial^{+}$ or $vestigial^{21}$ were subjected to DNA sequencing (see **Sequencing of PCR Products**) and the nature of the *vestigial* phenotype was evaluated.

DNA amplifications: DNA extracts from single flies were generated in a manner similar to the method described by GLOOR *et al.* (1993). However, in some instances the crude DNA preparations were further extracted with equal volumes of phenol/ chloroform/ isoamyl alcohol, precipitated with 3 volumes of 95% ethanol and resuspended in distilled water. The primers used in DNA amplification flank the *P* element insertion of $vestigial^{21}$ located at the 5' end of the gene and are described in HESLIP *et al.* (1992) as primer 1, 5'-AATCAAGTGGGCGGTGCTTG-3' and primer 7, 5'-ATCCCGCGCGGCGGTGAGAG-3'. The amplification reactions were carried out in a total volume of 30 ul starting with 3 ul of the above DNA extract (approximately 50 ng of genomic DNA) and a final concentration of 50 mM Tris-Cl (pH 9.2) 1.5 mM $MgCl_2$, 0.01% dithiothreitol , 0.1 ug/ul bovine serum albumin (Boehringer Mannheim), 200uM of dATP, dCTP, dGTP and dTTP, 0.3 ng/ul of primer and 1U of *Taq* DNA polymerase (BRL). The amplifications were produced utilizing a Stratagene Robocylcer 40 with a program of 1 cycle of 5 min at 95°C 1 min at 60°C and 3 min at 73°C, followed by 30 cycles of 1 min at 92°C 1 min at 60° C and 3 min at 73°C. Products of the amplification were fractionated by electrophoresis on 1 to 2% agarose gels in Tris-actate-EDTA. Events that did not produce an amplification product were not further analyzed.

Sequencing of PCR products: Some DNA sequencing was performed using a Sequenase II DNA sequencing kit as follows. Initially, PCR products amplified from several alleles were cloned into pT7T3-19U and subjected to double stranded plasmid sequencing according to manufacturer's instructions (New England Biolabs). PCR products were directly sequenced in the following manner. Approximately 0.3 to 0.5 ug of amplified DNA isolated utilizing the GenecleanII procedure from 1-2% agarose in TAE, was mixed with 100ng of primer, both dissolved in water, boiled for 8 minutes and immediately plunged into liquid nitrogen until ready to carry out the sequencing reactions. In other cases, DNA amplification products were purified from agarose gels with GenecleanII and sequenced using the ABI Taq DyeDeoxy Terminator Cycle Sequencing Kit according to the manufacturer's instructions. The sequencing reactions were fractionated and analyzed on an ABI 370A DNA sequencing apparatus.

RESULTS

In the first set of experiments, strong *vestigial* alleles were generated from the cryptic *vestigial*^{21} allele as described in MATERIALS AND METHODS. With the exception of targeted events, all of the strong *vestigial* alleles that were recovered contained deletions. In one screen 16 independent alleles were isolated from approximately 50,000 chromosomes scored: 12 had internal deletions, 3 had excisions of regions outside of the *P* element (and removed an adjacent *Eco*RI restriction site) and one was a targeted event (STAVELEY *et al.* 1994). All of the internally-deleted *vestigial* alleles share one striking feature: a breakpoint within the 31 base pair terminal inverted-repeat at the 3' end (TABLE 3.1). To differentiate these from other non-autonomous *P* elements, we refer to deletions which have breakpoints within one or both inverted-repeats but which retain the extreme terminal sequences as terminal internal deletions. The number of base pairs retained at the 3' end of these alleles ranges from 9 to 19. Most (15/17) retained the terminal 15 to 19 base pairs of the inverted-repeat. In contrast, the position of the 5' breakpoint of these internal deletions varies from nucleotide 105 (*vestigial*$^{21\text{-}1.3}$) to nucleotide 2714 (*vestigial*$^{21\text{-}6.1}$). Most of the 5' breakpoints are clustered in three regions of the *vestigial*^{21} *P* element: nucleotides 105 to 119, 142 to 148 and 2697 to 2714 (numbered according to the published sequence of the complete *P* element ; O'HARE and RUBIN 1983). These are located at the 5' side of very AT rich regions of the *P* element: the sequence from position 105 to 130 is 5'-GACGAATTTTTTTTTGAAAACATTAA-3' (80% AT) 142 to 171 is 5'-AATAAAAAAAAATGAAATATTGCAAATTTT-3' (90% AT) and 2686 to 2719 is 5'-AATTATTAAAAATAAAACTTTAAAAATAATTT-3' (97% AT). Alleles *vestigial*$^{21\text{-}1.3}$. *vestigial*$^{21\text{-}161.1}$ and *vestigial*$^{21\text{-}213.1}$ are identical but were isolated at different times and must have arisen independently. Apparently, deletions of the 3' region of the *P* element result in loss of the cryptic phenotype of *vestigial*^{21}.

TABLE 3.1

Analysis of deletion breakpoints in derivatives of vg^{21} selected for strong *vestigial* phenotype

allele	5' break	3' break	sequence of breakpoint junction	phenotype
vg^{21}	**380**	**2602(306)**	**...acctagtaaa/aaattagaat...**	**cryptic**
$vg^{21-1.3}$	105	2892(16)	...gttgtgtgcgg/atgttatttcatcatg	strap
$vg^{21-6.1}$	2714	2892(16)	...ctttaaaaata/atgttatttcatcatg	strap
$vg^{21-33.1}$	294	2889(19)	...tgatacccact/cttatgttatttcatcatg	extreme
$vg^{21-56.1}$	110	2891(17)	...gtgcggacgaa/tatgttatttcatcatg	extreme
$vg^{21-60.1}$	118	2891(17)	...gaattttttt/tatgttatttcatcatg	extreme
$vg^{21-74.2}$	2703	2892(16)	...taaaaataaaa/atgttatttcatcatg	strap
$vg^{21-78.1}$	148	2891(17)	...gtggaataaaa/TGTTTG/tatgttatttcatcatg	extreme
$vg^{21-81.1}$	142	2892(16)	...ccttacgtgga/atgttatttatcatg (1)	extreme
$vg^{21-83-1}$	130	2892(16)	...gaaacattaa/CATTAACATTAACATTAACATTA/atgttatttcatcatg	strap
$vg^{21-158.1}$	142	2892(16)	...ccttacgtgga/TAGTAAAATGTTTGT/atgttatttcatcatg	extreme
$vg^{21-161.1}$	105	2892(16)	...gttgtgtgcgg/atgttatttcatcatg	strap
$vg^{21-197.5}$	148	2891(17)	...gtggaataaaa/TGTTTG/tatgttatttcatcatg	extreme
$vg^{21-212.2}$	251	2899(9)	...agcagagcctt/ttcatcatg	extreme
$vg^{21-213.1}$	105	2892(16)	...gttgtgtgcgg/atgttatttcatcatg	strap
$vg^{21-213.4}$	2697	2892(16)	...aattattaaaa/atgttatttcatcatg	strap
vg^{21-241}	205	2897(11)	...aaattaattca/GTTAATTA/atttcatcatg	extreme
$vg^{21-f106}$	119	2893(15)	...aattttttttt/tgttatttcatcatg	extreme

(1) c at position 2901 is not present

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983) although the *P* element of *vestigial*^{21} is internally deleted (687 base pairs; WILLIAMS *et al.* 1988). The numbers in the 5' and 3' break columns indicate the nucleotide positions at either side of the internal deletion. The parenthetical number following the 3' breakpoint indicates the number of base pairs of the *P* element retained at this end. The DNA sequence of the junction is presented such that a slash (/) separates sequences assigned to either side of the breakpoint. Lowercase letters represent known *P* element sequences and uppercase represent filler sequences present at the breakpoint in a number of alleles.

Although 12 of the 17 events examined are simple deletions, the other five have AT rich filler sequences included at the breakpoint (TABLE 3.1). The *vestigial*$^{21\text{-}83.1}$ allele has four tandem copies of the hexameric repeat of 5'-CATTAA-3' as filler. Alleles *vestigial*$^{21\text{-}78.1}$ and *vestigial*$^{21\text{-}197.5}$ appear to be identical and have the sequence 5'-TGTTTG-3' at the breakpoint. These alleles were isolated from different dysgenic fathers and may have resulted from either two independent events or two separate isolations of a pre-existing variant in the *P[lacZ;ry*$^{+}$*]*^{c49}; *b cn vg*^{21}; *ry* stock. In support of the former, *vestigial*$^{21\text{-}81.1}$ and *vestigial*$^{21\text{-}7}$ (WILLIAMS *et al.* 1988) which were isolated under dissimilar experimental conditions separated by several years, share exactly the same deletion breakpoints and apparently only differ by the absence of the C at position 2901 in the *vestigial*$^{21\text{-}81.1}$ allele.

In the second screen, a series of phenotypic revertant alleles was generated in which the scalloped to strapped wing phenotype of *vestigial*^{21} hemizygotes was reverted to wild type. Revertants represented approximately 3 to 5 % of progeny. Two classes of internally-deleted *P* elements were isolated (TABLE 3.2). Eleven alleles were isolated which had breakpoints in both inverted-repeats and three alleles were isolated which had a breakpoint in the 5' inverted-repeat and another elsewhere within the *P* element. The positions of the 5' breakpoints range from the 13th to the 19th nucleotide. Most deletions (12 of 14) retain either 16 or 17 base pairs of the 5' inverted-repeat. Of the eleven revertant alleles that also have 3' breakpoints in the inverted-repeat 11 through 18 base pairs are retained and nine possess 16±1 base pairs of the terminus. Deletions of the 5' region of the *P* element result in a revertant phenotype as does deletion of both the 5' and 3' regions of the *P* element. The extreme alleles, by contrast, all retain at least approximately 100 base pairs of the 5' region of the *P* element but are missing 3' *P* element sequences.

Of the fourteen revertant alleles, seven possess filler sequences at the breakpoints. As with the strong *vestigial* alleles, these filler sequences are also AT rich. Most interesting is the inclusion of short tandem repeats of simple sequence in several of the mutants. For example, *vestigial*$^{21\text{-}IIIR1}$ and *vestigial*$^{21\text{-}IIIR5}$ have repeats of 5'-AACAT-3' and *vestigial*$^{21\text{-}IR6}$ has repeats of the complement 5'-ATGTT-3'. Other filler sequences, while not composed of simple tandem repeats, do have a somewhat repetitive structure such as the filler sequence (5'-TAATATATGTATATATTATAAATTATA-3') of *vestigial*$^{21\text{-}IR2}$. The sequences often appear to consist of tandem combinations of sequences present in the inverted-repeat. Exact assignment of the breakpoint in some

TABLE 3.2

Analysis of deletion breakpoints in derivatives of vg^{21} selected for reversion

allele	5' break	3' break	sequence of breakpoint junction	phenotype
vg^{21}	**380**	**2602(306)**	**...acctagtaaa/aaattagaat...**	**cryptic**
$vg^{21\text{-}IR2}$	17	2892(16)	catgatgaaataacata/TAATATATGTATATATTATAAATTATA/atgttatttcatcatg	revertant
$vg^{21\text{-}IR3}$	17	2891(17)	catgatgaaataacata/TA/tatgttatttcatcatg	revertant
$vg^{21\text{-}IR6}$	16	2890(18)	catgatgaaataacat/CATGTTATGTTATGTTATG/ttatgttatttcatcatg	revertant
$vg^{21\text{-}IR7}$	16	2892(16)	catgatgaaataac at gttatttcatcatg	revertant
$vg^{21\text{-}IR8}$	17	352(334)	catgatgaaataacata/ACTCAAT/gactcaacgca...	revertant
$vg^{21\text{-}IIR2}$	16	2895(13)	catgatgaaataaca t tatttcatcatg	revertant
$vg^{21\text{-}IIR8}$	16	2893(15)	catgatgaaataaca t gttatttcatcatg	revertant
$vg^{21\text{-}IIIR1}$	17	2893(15)	catgatgaaataacata/TATGTAACATAACATAACATAACATC/tgttatttcatcatg	revertant
$vg^{21\text{-}IIIR2}$	16	2897(11)	catgatgaaataac at ttcatcatg	revertant
$vg^{21\text{-}IIIR3}$	16	2836(72)	catgatgaaataacat/tgcacttattt...	revertant
$vg^{21\text{-}IIIR4}$	19	2859(49)	catgatgaaataaca taag tggatgtctc...	revertant
$vg^{21\text{-}IIIR5}$	17	2891(17)	catgatgaaataacata/TATGTAACATAACATAACATAACA/tatgttatttcatcatg	revertant
$vg^{21\text{-}IIIR7}$	13	2891(17)	catgatgaaataa/TAATGATAA/tatgttatttcatcatg	revertant
$vg^{21\text{-}IIIR8}$	16	2892(16)	catgatgaaataac at gttatttcatcatg	revertant

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983) and the sequence of the junctions are presented as in TABLE 3.1. Underlined nucleotides represents sequence that may have arisen from either side of the breakpoint.

alleles is impossible due to short regions of sequence identity (1, 2 or 4 base pairs) on either side of the breakpoint. For example *vestigial*$^{21\text{-}IIIR8}$ could have retained 16 base pairs of the 5' end or 16 at the 3' end but the entire length is only 30 base pairs.

Our observation that the majority of the deletion breakpoints are nested within the terminal inverted-repeats led us to question the existence of a bias inherent in the selection schemes. To examine this, derivatives of *vestigial*^{21} were isolated solely by alterations in the size of the *P* element. Five alleles were isolated which produced PCR-generated fragments of altered length (TABLE 3.3). All five alleles had breakpoints in both the 5' and 3' inverted-repeats and, therefore, resemble the majority of the alleles generated in our revertant screen. The 5' breakpoints range from bases 15 to 17 and the 3' breakpoints vary from 10 to 17 base pairs retained. All are phenotypic revertants. Although loss of only the 3' region of the *P* element results in strong *vestigial* alleles and loss of 5' *P* element sequences leads to phenotypic reversion, the positioning of the breakpoint within the inverted-repeat at approximately the 16th base pair from the end does not seem to result from a bias based upon phenotypic selection.

An examination of the distribution of the breakpoints that occurred within the inverted-repeat from our experiments is presented in FIGURE 3.1. We have also included data from the sequences of internally-deleted *P* element derived from an insertion of pUC*hsneo* at position 3C (*Xneo3* ; TAKASU-ISHIKAWA *et al*. 1992). Analysis of the distribution of breakpoints demonstrates that a very high percentage of the disrupted *P* element termini (70% combined) have a breakpoint at 15 to 17 base pairs from an end. The conservation of the terminal 16 base pairs of the *P* element suggests that these sequences are protected during the process that leads to terminal internal deletion. A very good candidiate for this activity is the inverted-repeat-binding protein (IRBP) isolated by RIO and RUBIN (1988).

DISCUSSION

These experiments amount to an *in vivo* analysis of transposase mediated deletions of the *P* element of *vestigial*^{21} (FIGURE 3.2). Deletions of the 3' region of the *P* element, as small as the 177 basepair deletion between positions 2714 and 2892 of *vestigial*$^{21\text{-}6.1}$ result in a strong *vestigial* phenotype. Deletions of the 5' region of the *P* element, as well as deletions that remove everything but the very 5' and 3' termini, result in revertant phenotypes. The site where polyadenylation of the *P* element transcript

TABLE 3.3

Analysis of deletion breakpoints in derivatives of vg^{21} selected for altered PCR fragment length

allele	5' break	3' break	sequence of breakpoint junction	phenotype
vg^{21}	**380**	**2602(306)**	**...acctagtaaa/aaattagaat...**	**cryptic**
vg^{21}*-G3.11*	15	2895(13)	catgatgaaataaca/ttatttcatcatg	revertant
vg^{21}*-G4C6*	16	2898(10)	catgatgaaataacat/tttcatcatg	revertant
vg^{21}*-G5B1*	16	2892(16)	catgatgaaataac at gttatttcatcatg	revertant
vg^{21}*-G6B3*	17	2891(17)	catgatgaaataaca ta tgttatttcatcatg	revertant
vg^{21}*-G7B9*	16	2895(13)	catgatgaaataacat/ttatttcatcatg	revertant

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983) and the sequence of the junctions are presented as in TABLE 3.1 and TABLE 3.2.

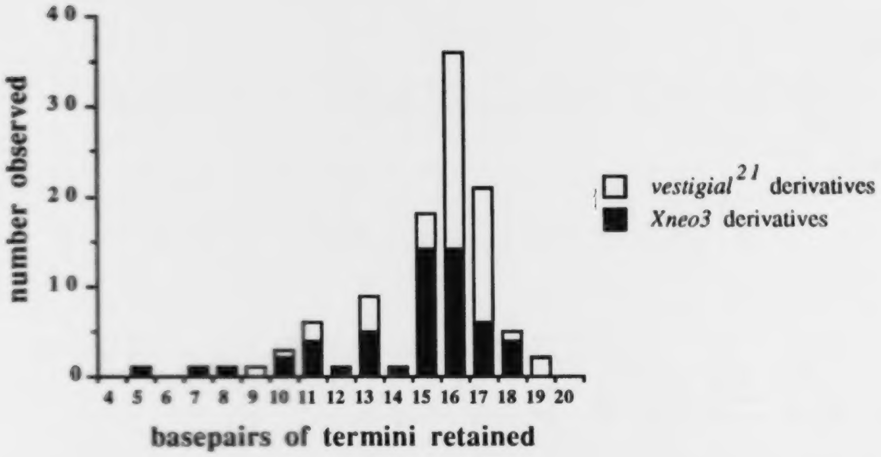
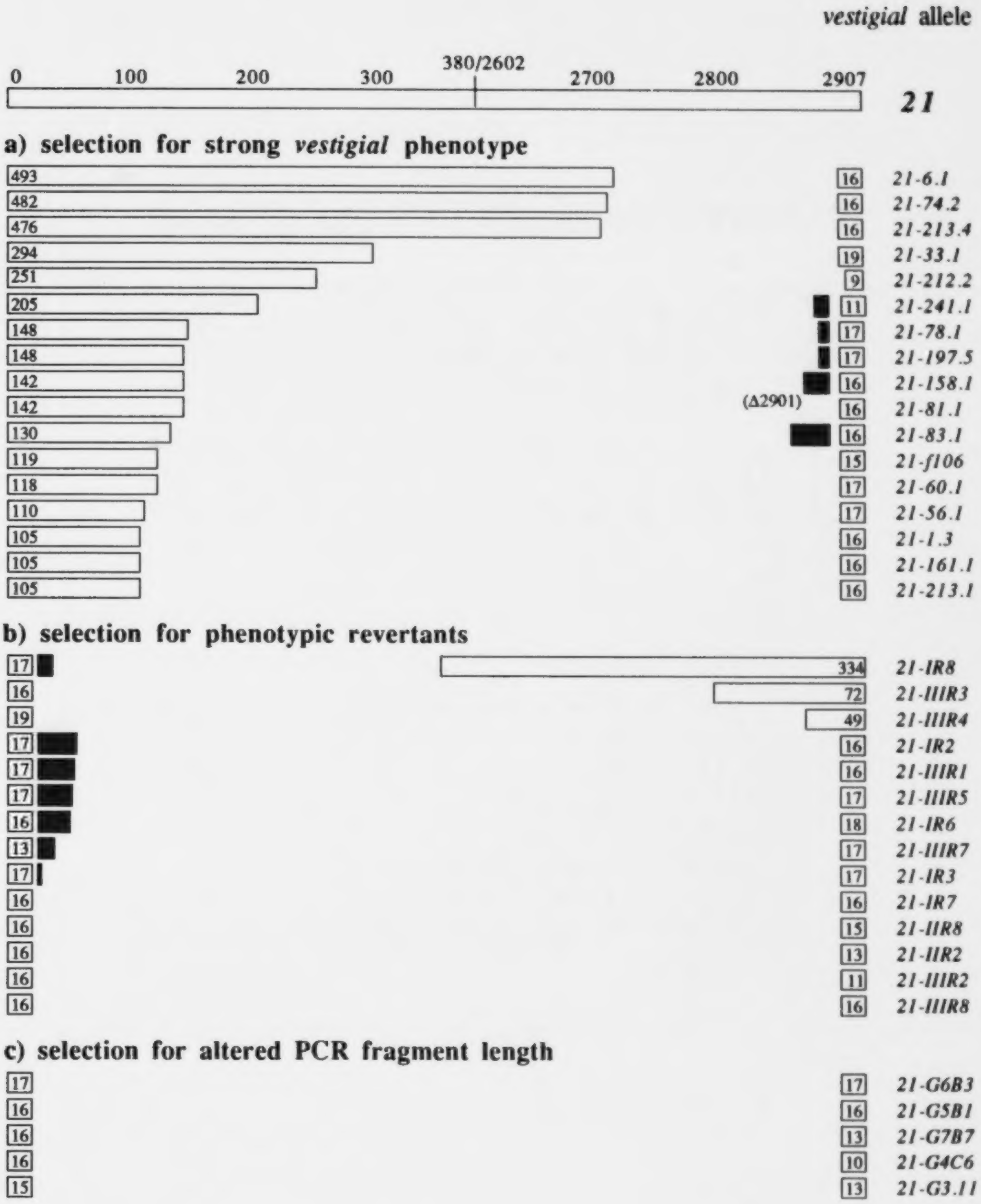


FIGURE 3.1. Distribution of the breakpoint position in terminal internal deletions. The number of *P* element terminal sequences (5 to 19 base pairs) of both *vestigial*^{21} (this study) and *Xneo3* derivatives (TAKASU-ISHIKAWA *et al.* 1992) is presented. Note the strong tendency to retain 16±1 base pairs of the terminus. It should be noted that the position of the breakpoints has been assignedto the site where the sequence diverges from that of the terminal inverted repeat. The inclusion of ambiguous residues would bias the data slightly to the right.

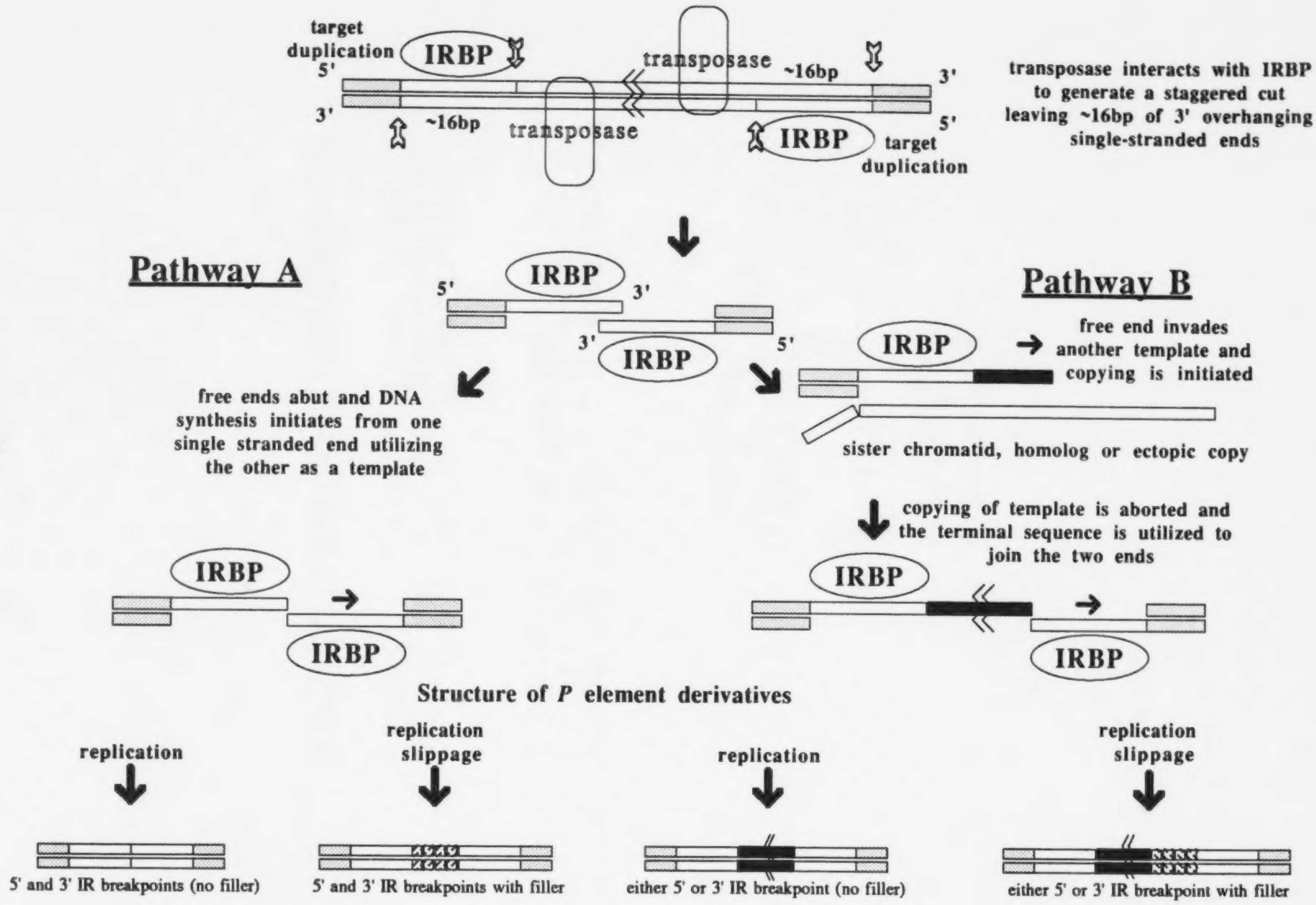
FIGURE 3.2. Terminal internal deletion derivatives of *vestigial*^{21} . The open box at the top represents the sequence of the *P* element located in the 5' region of the transcription unit in the *vg*^{21} allele. The line in the centre of this box indicates the internal deletion breakpoint of this *P* element. The open boxes below represent the *P* element sequences present in the allele indicated at the right. The numbers within the open boxes correspond to the number of base pairs of *P* element sequences, according to the published sequence (O'HARE AND RUBIN 1983), remaining on that side of the breakpoint. The areas between the open boxes represents internally-deleted *P* element sequences. The black boxes represent filler sequences located at the deletion breakpoints.


vestigial allele
0
100
200
300
380/2602
2700
2800
2907
21
a) selection for strong *vestigial* phenotype
493 16 *21-6.1*
482 16 *21-74.2*
476 16 *21-213.4*
294 19 *21-33.1*
251 9 *21-212.2*
205 11 *21-241.1*
148 17 *21-78.1*
148 17 *21-197.5*
142 16 *21-158.1*
142 (Δ2901) 16 *21-81.1*
130 16 *21-83.1*
119 15 *21-f106*
118 17 *21-60.1*
110 17 *21-56.1*
105 16 *21-1.3*
105 16 *21-161.1*
105 16 *21-213.1*
b) selection for phenotypic revertants
17 334 *21-IR8*
16 72 *21-IIIR3*
19 49 *21-IIIR4*
17 16 *21-IR2*
17 16 *21-IIIR1*
17 17 *21-IIIR5*
16 18 *21-IR6*
13 17 *21-IIIR7*
17 17 *21-IR3*
16 16 *21-IR7*
16 15 *21-IIR8*
16 13 *21-IIR2*
16 11 *21-IIIR2*
16 16 *21-IIIR8*
c) selection for altered PCR fragment length
17 17 *21-G6B3*
16 16 *21-G5B1*
16 13 *21-G7B7*
16 10 *21-G4C6*
15 13 *21-G3.11*

occurs (LASKI *et al.* 1986) is absent in all the extreme alleles generated. All of the phenotypic revertants have 5' deletions (with a breakpoint at ~16 bp of the 5' inverted-repeat) and all the strong *vestigial* alleles retain 5' sequences to at least the 105th base pair (ie. *vestigial*$^{21\text{-}1.3}$). The region between the 5' inverted-repeat and position 105 contains both the promoter and the transcriptional start site of the *P* element transcription unit (KARESS and RUBIN 1984). We suggest that transcription initiated at the *P* element promoter which is not terminated within the *P* element interferes with *vestigial* gene function. The termination of transcription within the *P* element, as in *vestigial*^{21}, only moderately effects *vestigial* expression and results in the cryptic phenotype. Removal of the *P* transcriptional start site seems to return *vestigial* expression to wild type levels even if several hundred base pairs of the 3' end of the *P* element, as in *vestigial*$^{21\text{-}IR8}$, remain at the insertion site.

The tendency for deletion derivatives of *P* element alleles in *Drosophila melanogaster* to have an endpoint within the 31 base pair terminal inverted-repeat has been reported previously (SEARLES *et al.* 1986; WILLIAMS *et al.* 1988; JOHNSON-SCHLITZ and ENGELS 1993) and is evident in the extensive data set of TAKASU-ISHIKAWA *et al.* (1992). We wish to propose a model, a derivative of the synthesis-dependent strand annealing (SDSA) model (NASSIF et. al. 1994) to explain the formation of terminal internal deletions (FIGURE 3.3). The transposase, which binds at sites just inside the inverted-repeats (KAUFMAN *et al.* 1989), must cut at both ends of the transposon (5'-CATG-3') to allow transposition. To satisfy the requirements of the cut-and-paste transposition model to generate target site duplications (SHAPIRO 1979; KAUFMAN and RIO 1992) and assuming that excision is an initial event in a procedure which may or may not be followed by integration, a mechanism, which generates either 3' overhanging or blunt ends, must exist. A 16 nucleotide 3' overhang could be specified by the binding of the IRBP through interaction with the transposase. A staggered cut of this length is much longer than the 4 to 7 nucleotide overhang which the recovery of plasmid borne excision products from injected embryos would seem to suggest (O'BROCHTA *et al.* 1991). However, *P* elements excised from plasmids transiently placed in the embryonic soma may not be subjected to the same constraints as *P* elements integrated into the germline. We propose that the IRBP has a role in the stabilization of the 16 nucleotide 3' overhang and thus participates in repair of the double-strand break. In the case of the very smallest *P* elements recovered, those that retain only 16 or so base pairs of each terminus fused together, the simplest model to explain their origin involves the two 3' overhanging ends abutting or overlapping by a basepair or two, followed by DNA

FIGURE 3.3. Possible role for IRBP in transposase-induced double-strand gap repair. The sequence analysis of *P* element terminal internal deletions suggest that the inverted-repeat-binding protein (IRBP) may maintain the terminal ~16 base pairs of the *P* element after excision of the *P* element at the donor site. The protected sequences can then be joined by a synthesis-dependent reaction by which DNA synthesis results in bridging the gap between free ends (Pathway A). Invasion of another template (Pathway B) prior to this results in either recovery of at least some of the internal *P* element sequences (if an identical *P* element is copied) or of different *P* element sequences - targeted transposition- (if another *P* element is copied). Generation of filler sequences can be explained by rounds of replication and slippage during the synthesis-dependent joining of the ends. The diagrams of Pathway B account for the 3' *P* element deletions of FIGURE 3.2a. Deletions of the 5' *P* element sequences would occur when the 3' terminus invades another template.


target duplication
5'
3'
IRBP
~16bp
transposase
transposase
~16bp
3'
5'
IRBP
target duplication
transposase interacts with IRBP to generate a staggered cut leaving ~16bp of 3' overhanging single-stranded ends
Pathway A
5'
IRBP
3'
3'
IRBP
5'
Pathway B
free ends abut and DNA synthesis initiates from one single stranded end utilizing the other as a template
IRBP
free end invades another template and copying is initiated
sister chromatid, homolog or ectopic copy
copying of template is aborted and the terminal sequence is utilized to join the two ends
IRBP
IRBP
IRBP
IRBP
Structure of *P* element derivatives
replication
replication slippage
replication
replication slippage
5' and 3' IR breakpoints (no filler)
5' and 3' IR breakpoints with filler
either 5' or 3' IR breakpoint (no filler)
either 5' or 3' IR breakpoint with filler

synthesis to cross the gap (FIGURE 3.3; pathway A). The IRBP has the potential to mediate this activity by maintaining the single-stranded 3' overhang thus preserving a template to be copied. Larger *P* elements which have a breakpoint in only one of the inverted-repeats seem to have had one end invade a homologous sequence (FIGURE 3.3; pathway B). However, before the entire element is copied, DNA synthesis switches to copy the terminal sequences. The template switching often seems to occur in the 5' regions of AT rich sequences. Either end can initiate synthesis since we have isolated terminal internal deletions which break in either the 5' or the 3' repeat. This can also explain the recovery, at *white*, of truncated *yellow* sequences fused to 14 or 17 base pairs of the *P* element terminus (NASSIF *et al.* 1994). In addition, the recovery of chimeric *P* elements in which flanking DNA sequences were recruited from a donor site (TSUBOTA and DANG-VU 1991; HESLIP *et al.* 1992) may be explained in a similiar manner, except that DNA synthesis has apparently continued through the ectopic *P* element and into adjacent sequences (NASSIF *et al.* 1994) and finally resolved by switching to the inverted-repeat template.

An examination of the filler sequences included at the breakpoint reveals an interesting property. Short tandem repeats of simple sequence of 5'-AACAT-3', 5'-ATGTT-3' or 5'-TAACAT-3' are present in some alleles. If we compare these sequences to the first 18 base pairs of the terminal inverted-repeat (5'-catgatgaaataacataa-3'), the repeats are identical to sequences at positions 10 through 18 (or its complement). This observation suggests that, at least in some cases of filler sequences found at the deletion breakpoint, a mechanism exists to copy at this junction short sequences identical to regions of the inverted-repeat. For example, the filler sequence from *vestigial*$^{21\text{-}78.1}$ is 5'-TGTTTGT-3' and could represent bases 2893 through 2896 (5'-TGTT-3') and 2893 through 2895 (5'-TGT-3') joined in tandem. In the specific case of the *vestigial*$^{21\text{-}83.1}$ allele, the 5'-CATTAA-3' repeat could have been generated either as described above or through a similar mechanism copying a sequence located between position 123 and 131 (5'-AACATTAAC-3'). A similar doublet 5'-ATCATTA-3', found in the cloned *P* element λπ51 (π13B) (O'HARE and RUBIN 1983; O'HARE *et al.* 1992), could also have been copied from the adjacent 5'-ATCATTATC-3' between positions 2114 and 2122. In both cases, the alternate repeats are very similar and may have arisen under similar circumstances. To generate filler sequences, we favour a model based on the process of replication-slippage-replication originally proposed by O'HARE and RUBIN (1983) and supported by KURKULOS *et al.* (1994). This model is very different from the hairpin model proposed by TAKASU-ISHIKAWA *et al.* (1992). The generation of the short tandem repeats of simple sequence in *vestigial*$^{21\text{-}IIIR1}$, *vestigial*$^{21\text{-}IIIR5}$, *vestigial*$^{21\text{-}IR6}$ and *vestigial*$^{21\text{-}83.1}$ can readily be attributed to several rounds of replication, slippage and replication utilizing part of the inverted-repeat sequence (5'-catgatgaaataacataa-3' or a similar sequence) or its complement as a template. Most of the other filler sequences are highly AT rich and some of the less structured filler sequences could have arisen in a similar manner. The region of the inverted-repeat which appears to be utilized as template in these proposed rounds of replication and slippage correspond to the proximal sequences protected by the IRBP in the DNA protection assays (RIO and RUBIN 1988).

The inverted-repeat-binding protein (IRBP) is a host derived protein and its normal activity is unclear. One correlation, which may have some biological significance, should be mentioned. In studies of *Drosophila melanogaster* satellite sequences, LOHE and BRUTLAG (1987) described stretches of short repetitive sequences that form a smooth transition from one repeat to another without losing the periodicity of the repeated structure. For example, a series of $(AAXAY)^n$ and $(AAXAZ)^m$ would show the junction as ...AAXAY AAXAY AAXAZ AAXAZ... (X, Y and Z may represent A,C,G or T). They proposed a model of template amplification to explain the origin of smooth junctions between two different but similar satellite sequences. According to this model, during the synthesis of an array of satellite DNA repeats, each repeat is generated using the preceding unit as a template. Precise replication results in long homogeneous tandem arrays while errors in replication would result in misreading of the repeat length or the introduction of a base change in the array followed by perpetuation of the error. This mechanism would produce the smooth junctions observed. The repeats observed in some of the filler sequences resemble satellite sequences, 5'-AACAT-3' fits the general structure of most pentameric tandem repeats of 5'-AAXAY-3' but not one previously described (LOHE *et al.* 1993). If the IRBP has a role in the generation of filler sequences, as we propose, then perhaps the protein is also involved in processes such as template amplification.

Recently, the IRBP has been identified as the Drosophila homolog (RIO, reported by PLASTERK 1993; BEALL *et al.* 1994) of the 70-kDa subunit of the human Ku autoimmune antigen, a DNA-binding protein (REEVES and STHOEGER 1989). If the two proteins share similar activities, it is possible that the Ku antigen may be also involved in the maintenance of some microsatellite sequences. The instability of particular microsatellites is believed to result in some genetic diseases including fragile X syndrome, Kennedy's disease, myotonic dystrophy, Huntington's disease and spinocerebellar ataxia type 1 (KUNKEL 1993). Such instability is believed to result from strand slippage during DNA synthesis.

LITERATURE CITED

BEALL, E .L., A. ADMON and D. C. RIO, 1994 A Drosophila protein homologous to the human p70 Ku autoimmune antigen interacts with the *P* transposable element inverted repeats. Proc. Natl. Acad. Sci. USA **91**: 12681-12685.

ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 437-484 in *Mobile DNA.*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington DC.

ENGELS, W. R., D. M. JOHNSON-SCHLITZ, W. B. EGGLESTON and J. SVED, 1990 High-frequency *P* element loss in Drosophila is homolog dependent. Cell **62**: 515-525.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. R. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in Drosophila via *P* element-induced gap repair. Science **253**: 1110-1117.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS, W. K. BENZ, H. M. ROBERTSON and W. R. ENGELS, 1993 Type I repressors of *P* element mobility. Genetics **135**: 81-95.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131**: 917-927.

HESLIP, T. R., and R. B. HODGETTS, 1994 Targeted transposition at the *vestigial* locus of *Drosophila melanogaster*. Genetics **138**: 1-9.

JOHNSON-SCHLITZ, D. M., and W. R. ENGELS, 1993 *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila* melanogaster. Mol. Cell Biol. **13**: 7006-7018.

KARESS, P. D., and G. M. RUBIN, 1984 Analysis of *P* transposable element functions in *Drosophila*.. Cell **38**: 135-146.

KAUFMAN, R. E., R. F. DOLL and D. C. RIO, 1989 Drosophila *P* element transposase recognizes internal *P* element DNA sequences. Cell **59**: 359-371.

KAUFMAN, R. E., and D. C. RIO, 1992 *P* element transposition *in vitro* proceeds by a cut and paste mechanism and uses GTP as a cofactor. Cell **69**: 27-39.

KUNKLE, T. A., 1993 Slippery DNA and diseases. Nature **365**: 207-208.

KURKULOS, M., J. M. WEINBERG, D. ROY and S. M. MOUNT, 1994 *P* element-mediated *in vivo* deletion analysis of *white-apricot*: deletions between direct repeats are strongly favored. Genetics **136**: 1001-1011.

LASKI, F. A., D. C. RIO, and G. M. RUBIN, 1986 Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44**: 7-19.

LASKO, P. F., and M. L. PARDUE, 1988 Studies of the genetic organization of the *vestigal* microregion of *Drosophila melanogaster*. Genetics **120**: 495-502.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

LOHE, A. R., and D. L. BRUTLAG, 1987 Adjacent satellite DNA segments in Drosophila. Structure of junctions. J. Mol. Biol. **194**:171-179.

LOHE, A. R., A. J. HILLIKER and P. A. ROBERTS, 1993 Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. Genetics **134**: 1149-1174.

MISRA, S., and D. C. RIO, 1990 Cytotype control of Drosophila *P* element transposition: the 66kD protein is a repressor of transposase activity. Cell **62**: 269-384.

MULLINS. M. C., D. C. RIO and G. M. RUBIN, 1989 *Cis*-acting DNA sequence requirements for *P* element transposition. Genes Dev. **3**: 729-738.

NASSIF, N., J. PENNEY, S. PAL, W. R. ENGELS and G. B. GLOOR, 1994 Efficient copying of nonhomologous sequences from ectopic site via *P*-element-induced gap repair. Mol. Cell Biol. **14**: 1613-1625.

O'BROCHTA, D. A., S. P. GOMEZ and A. M. HANDLER, 1991 *P* element excision in *Drosophila melanogaster* and related drosophilids. Mol. Gen. Genet. **225**: 387-394.

O'HARE, K., and G. M. RUBIN, 1983 Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**: 25-35.

O'HARE, K., A. DRIVER, S. MCGRATH and D. M. JOHNSON-SCHLITZ, 1992 Distribution and structure of cloned *P* elements from the *Drosophila melanogaster P* strain π_2. Genet. Res., Camb. **60**: 33-41.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84**: 9123-9127.

PLASTERK, R. H. A., 1993 Molecular mechanisms of transposition and its control. Cell **74**: 781-786.

REEVES, W. H., and Z. M. STHOEGER, 1989 Molecular cloning of cDNA encoding the p70 (Ku) lupus autoanitgen. J. Biol. Chem. **264**: 5047-5052.

RIO., D. C., 1990 Molecular mechanisms regulating Drosophila *P* element transposition. Ann. Rev. Genet. **24**: 543-578.

RIO, D. C., and G. M. RUBIN, 1988 Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element. Proc. Natl. Acad. Sci. USA **85**: 8929-8933.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ and W. R. ENGELS, 1988 A stable genetic source of *P*-element transposase in *Drosophila melanogaster*. Genetics **118**:461-470.

SEARLES L. L., A. L.GREENLEAF, W. E. KEMP and R. A. VOELKER, 1986 Sites of *P* element insertion and structures of *P* element deletions in the 5' region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6**: 3312-3319.

SHAPIRO, J. A., 1979 Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA **76**: 1933-1937.

STAVELEY, B. E., R. B. HODGETTS, S. L. O'KEEFE, and J. B. BELL, 1994 Targeting of an enhancer trap to *vestigial*. Developmental Biology **165**: 290-293.

TAKASU-ISHIKAWA, E., M. YOSHIHARA and Y. HOTTA, 1992 Extra sequences found at *P* element excision sites in *Drosophila melanogaster*. Mol. Gen. Genet. **232**: 17-23.

TSUBOTA, S. I., and H. DANG-VU, 1991 Capture of flanking DNA by a *P* element in *Drosophila melanogaster* : creation of a transposable element. Proc. Natl. Acad. Sci. USA **88**: 693-697.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**:1355-1363.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988 Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

Chapter 4

A model of transcriptional interference for *P* element alleles of *vestigial*

The phenomenon of *P-M* hybrid dysgenesis in *Drosophila melanogaster* is a syndrome of defects of the germline, including high mutability, male recombination and gonadal dysgenic sterility, that is associated with the mobilization of the *P* transposable element (reviewed in ENGELS 1989; RIO 1990; 1991). This syndrome occurs in the progeny of crosses between *P*-strain males and *M*-strain females but not in the reciprocal cross. The complete *P* element is 2907 base pairs in length (O'HARE and RUBIN 1983) and encodes both the *P* element transposase (KARESS and RUBIN 1984) and a somatic repressor (MISRA and RIO 1990). Transposition occurs through a cut-and-paste mechanism (KAUFMAN and RIO 1992) which requires the presence of a functional *P* element transposase gene (KARASS and RUBIN 1984) and the ends of the *P* element (MULLINS *et al.* 1989). Many *P*-element insertions are stable in *M*-strains, due to the absence of a transposase source, and are also immobile in *P*-strains due to the active repression of transposition in the presence of intact *P* elements.

P element transposition is repressed by products encoded by the *P* elements themselves. This circumstance is referred to as the *P* cytotype. The mechanism which directs the repression of *P* element activity has been addressed by several models. In one suggestion, the presence of *P* element ends has been proposed to bind the transposase and other factors involved in the transposition of *P* elements and thus prevent transposition (SIMMONS and BUCHOLZ 1985). Another model suggests that some *P* elements produce an antisense RNA species which could act to repress *P* element activity (RASMUSSON, *et al.* 1993). A very compelling model, which does not exclude the influence of the preceding proposals, involves the production of a polypeptide repressor (MISRA and RIO 1990). The 66 kDa somatic repressor is a truncated version of the transposase protein and results from an alternatively spliced version of the *P* element message. Several repressor mechanisms involving this protein have been proposed which include transcriptional regulation, binding-site competition and multimer poisoning (ENGELS 1989; RIO 1990; 1991). The transposase has been shown to repress transcription *in vitro* by binding to a site at the 5' end of the *P* element (KAUFMAN *et al.* 1989, KAUFMAN and RIO 1992) to prevent the binding of the transcription factor (TFIID) at an overlapping site. To produce a type I repressor, a message initiated at the *P* element promoter must be transcribed to or past a position 1953 (±3) nucleotides from the 5' end of the complete *P* element (GLOOR *et al.* 1993). Modified *P* elements which encode the 66 kDa repressor protein did not initially display maternal repression of *P* element activity as monitored by the *singed*^{weak} test (MISRA and RIO 1990). However, mobilization to several new positions did result in significant maternal repression (MISRA *et al.* 1993). Control of gene activity in the germline is only part of the influence that the products of the *P* element may have upon gene expression. Genes other than the repressor/transposase transcription unit that have been placed under the transcriptional control of *P* element sequences are presumably susceptible to, at the very least, a subset of the mechanisms that direct *P* element transcription.

Some *P* element alleles of certain genes have unique properties, one of which is a phenotypic sensitivity to the presence of *P* elements. It has been determined that several strong *P* element alleles of the ***singed*** locus are suppressed in the presence of active *P* elements (ROBERTSON and ENGELS 1989). A *P* element, modified *in vitro*, containing the *P[white*^{duplicated}*]* construct has been shown to be sensitive to the presence of *P* elements (COEN 1990). Studies of *P-lacZ* fusions have shown that *P* element regulatory products can inhibit the activity of the *P* promoter in somatic tissues (LEMAITRE and COEN 1991) and in the germline (LEMAITRE *et al.* 1993). The latter seems to be consistent with the processes that result in the maternal inheritance of the *P* cytotype. The control of the reporter gene outside of the germline apparently reflects the influence of cytotype upon expression of genes expressed in the soma. Through selection for extreme *vestigial* alleles from dysgenic crosses and the subsequent isolation from these of phenotypic revertants, deletions and recruitment derivatives of the *vestigial*^{21} allele were isolated (WILLIAMS *et al.* 1988a). Some of these alleles are sensitive to the presence of *P* elements and exhibit a cytotype-dependent phenotype (WILLIAMS *et al.* 1988b). This property has been exploited as a test for the presence of active *P* repressor proteins (GLOOR *et al.* 1993). Recently, we have examined a larger set of internal deletions that has given us insights into the mechanisms by which *P* elements excise and repair (STAVELEY *et al.* 1995). In addition, it is suggested that transcription initiated in the *P* element interferes with *vestigial* expression. Here, this suggestion is expanded and evidence is presented that a *P* element *vestigial* fusion transcript is produced by *vestigial*^{21} and some of its derivatives. A model of transcriptional interference is proposed to account for the range of phenotypes observed among the *P* alleles of *vestigial* and the sensitivity of some alleles to the presence of *P* elements.

MATERIALS AND METHODS

Drosophila strains: The *b cn vg*^{21} strain possesses an internally-deleted 687 bp (positions 1-380/2602-2907 of the published sequence; O'HARE and RUBIN 1983) *P* element inserted in the 5' untranscribed region of *vestigial* (WILLIAMS and BELL 1988;

WILLIAMS *et al.* 1988a). A number of *vestigial*^{21} derivatives have been described previously (WILLIAMS *et al.* 1988a; b; STAVELEY *et al.* 1995) and many extreme *vestigial* alleles were generated during experiments to study targeted transposition (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994; STAVELEY *et al.* 1994). In general, *b cn vg*^{21} individuals, often possessing other modified *P* elements, were crossed to a stable source of transposase provided by the *w; Sb P[Δ2-3,ry*$^{+}$*] e/TM6, e* stock (ROBERTSON *et al.* 1988). Dysgenic males were mated *b cn vg*^{21} females to select for strong *vestigial* alleles. Individuals heterozygous for *vestigial*^{21} and a strong *vestigial* allele have nicked or scalloped wings but *vestigial*^{21} homozygotes display a wild type phenotype. Offspring displaying any wing-margin defects were crossed to establish lines. Revertants of *vestigial*^{21} (the scalloped to strapped wing phenotype observed in *vestigial*^{21} hemizygotes) were generated by screening the progeny of dysgenic males of the genetic constitution *vg*^{21}*/+; Sb P[Δ2-3,ry*$^{+}$*] e/+* through crosses to *Df(2R)vg*^{B}*/CyO* females (LASKO and PARDUE 1989). Other Drosophila mutations are described in LINDSLEY and ZIMM (1992). All flies were maintained at 22°C on a standard yeast/glucose medium (NASH and BELL 1968).

Recombinant DNA technology: DNA manipulations were performed according to standard methods (SAMBROOK *et al.* 1989). The restriction map of the 45 kb genomic region including the *vestigial* gene (WILLIAMS and BELL. 1988) and the exon/intron map of the *vestigial* transcription unit (WILLIAMS *et al.* 1990a) have been previously reported.

RNA isolation: Total Drosophila RNA extracts were prepared as described by O'KEEFE *et al.* (1995). Approximately ten fresh white pupae were placed in a sterile, baked RNase-free glass homogenizer and 400 ul of RNA isolation buffer (0.1 M Tris, 0.1 M NaCl, 20 mM EDTA and 1% SDS) was added. One fast grinding stroke of the pestle (2 to 3 seconds) was quickly followed by the addition of 400 ul of phenol equilibrated with sterile DEPC-treated distilled water. The tissue was homogenized in the phenol/buffer mixture for ten strokes. The mixture, transferred to microfuge tubes, was separated into phases by a 30 second 12000 rpm centrifugation. Carefully avoiding contamination of material from the phenol/water interface, the aqueous phase was removed and quickly extracted three times with equal volumes of diethyl ether. Cold ethanol (2.5 volumes) was added to the extracted sample and precipitation was allowed to occur for 30 minutes at room temperature followed by a 5 minute microcentrifugation. The remaining liquid was removed by micropipette and the pellet was dried for 5 minutes under low pressure. This RNA preparation was resuspended in 20 ul of DEPC-treated water (2 ul per animal) and stored at -80°C until needed.

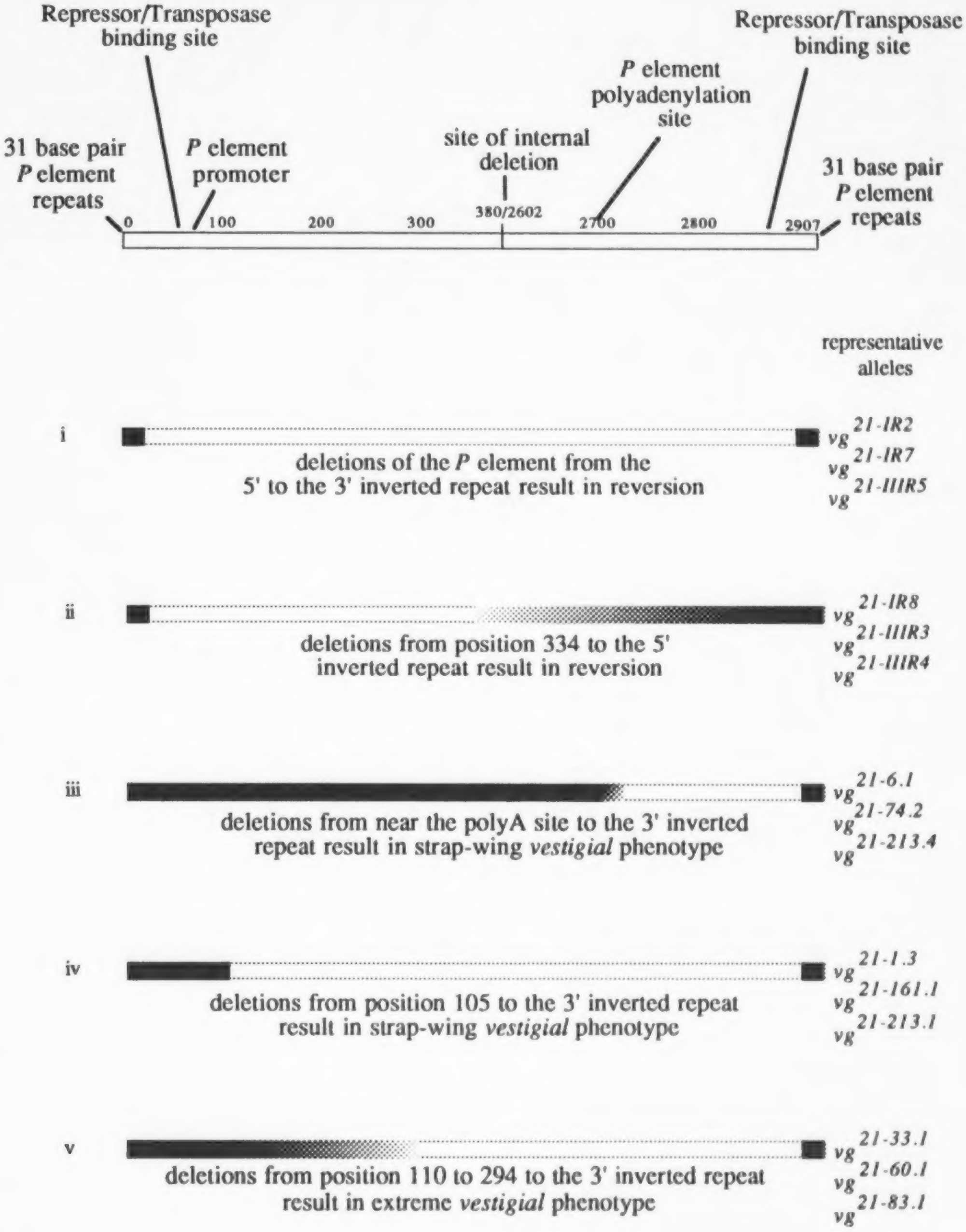
Northern analysis: Total RNA was isolated from *vg*^{21}, *vg*$^{21\text{-}6.1}$, *vg*$^{21\text{-}60.1}$, *vg*$^{21\text{-}1.3}$ and Oregan R then fractionated by electrophoresis according to standard methodologies (SAMBROOK *et al.* 1992). The RNA was transferred by capillary blotting to nitrocellulose membranes and hybridization was carried out at 65°C. The *Eco*RI-*Hinc*II fragment of *vg* cDNA2, which corresponds to sequences from exons 1, 2 and 3 of the 3.8 kb *vestigial* mRNA, was isolated by agarose gel electrophoresis, purified by GeneCleanII (Bio 101 Inc.) and oligo-labeled in the presence of ^{32}P dCTP (ICN) to generate a radioactive probe. The blot was washed under stringent conditions and exposed to X-ray film.

RESULTS

The initial step in the cloning of the *vestigial* gene involved the transposon tagging of the locus to produce the original *P* element allele, *vestigial*^{21} (WILLIAMS and BELL 1988). The molecular events which resulted in the secondary and tertiary derivatives of *vestigial*^{21} included both internal and adjacent deletions of the *P* element sequences of *vestigial*^{21} and recruitment of larger *P* elements to the original insertion site (WILLIAMS *et al.* 1988a). The latter property has been exploited to produce targeted transposition events at *vestigial* (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994; STAVELEY *et al.* 1994). Recently a relatively large series of internally-deleted derivatives of the *vestigial*^{21} *P* element has been described (STAVELEY *et al.* 1995). Although the focus of that work was to examine the tendency of the *P* element ends to remain at the donor site, these experiments amounted to an *in vivo* analysis of transposase mediated deletions of the *P* element of *vestigial*^{21} and are summarized in FIGURE 4.1.

Correlation of *vestigial* phenotype to *P* element lesion: The original *P* element allele of *vestigial*, *vestigial*^{21}, has an internally deleted 0.7 kb insert and displays a wild type phenotype. The deletion derivatives of *vestigial*^{21} fall into two main phenotypic categories: revertants (FIGURE 4.1 i and ii) and strong *vestigial* mutants (FIGURE 4.1 iii, iv and v). Among these mutants, the absence of the 3' region of the *vestigial*^{21} *P* element produces a strong *vestigial* phenotype (FIGURE 4.1 iii, iv and v). The smallest recovered deletion of this type, from the *vestigial*$^{21\text{-}6.1}$ allele, has 177 base pairs absent, when compared to the parental *vestigial*^{21} *P* element, between positions 2714 and 2892

FIGURE 4.1. The *P* element of *vestigial*^{21}. The *P* element features indicated above the diagram are taken from a number of studies (O'HARE and RUBIN, 1983; KARESS and RUBIN, 1984; WILLIAMS *et al.*, 1988a; KAUFMAN and RIO, 1991; 1992). The bars below represent different classes of internal deletion of the *P* element. The top two bars demonstrate that deletions of all but the ends and deletions of the 5' portions of the *P* element result in reversion of the *vestigial*^{21} phenotype. Strap-wing mutants arise when small 3' deletions or when large deletions with a breakpoint near the promoter are generated. Deletions of sequences from between position 110 and 294 to the 3' end produce a "classic" *vestigial* phenotype.

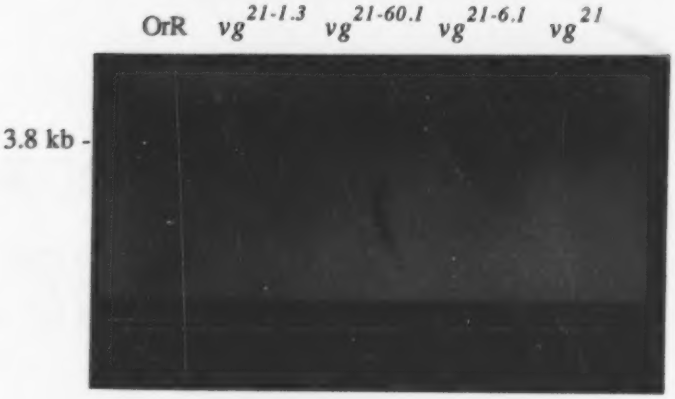

Repressor/Transposase binding site
Repressor/Transposase binding site
P element polyadenylation site
site of internal deletion
31 base pair P element repeats
P element promoter
31 base pair P element repeats
0
100
200
300
380/2602
2700
2800
2907
representative alleles
i
deletions of the P element from the 5' to the 3' inverted repeat result in reversion
vg 21-IR2
vg 21-IR7
vg 21-IIIR5
ii
deletions from position 334 to the 5' inverted repeat result in reversion
vg 21-IR8
vg 21-IIIR3
vg 21-IIIR4
iii
deletions from near the polyA site to the 3' inverted repeat result in strap-wing vestigial phenotype
vg 21-6.1
vg 21-74.2
vg 21-213.4
iv
deletions from position 105 to the 3' inverted repeat result in strap-wing vestigial phenotype
vg 21-1.3
vg 21-161.1
vg 21-213.1
v
deletions from position 110 to 294 to the 3' inverted repeat result in extreme vestigial phenotype
vg 21-33.1
vg 21-60.1
vg 21-83.1

of the published *P* element sequence. Deletions of the 5' region of the *vestigial*^{21} *P* element result in revertant phenotypes (FIGURE 4.1 i and ii). The revertant class includes large internal deletions that retain only part of each of the terminal repeats (FIGURE 4.1 i). All of the characterized phenotypic revertants have a breakpoint in the 5' inverted-repeat and another downstream within the *P* element. The absence of the *P* transcriptional start site appears to return *vestigial* expression to wild type levels even if several hundred base pairs of the 3' end of the *P* element remain at the insertion site as with the *vestigial*$^{21-IR8}$ allele (FIGURE 4.1 ii). In contrast, the strong *vestigial* alleles retain, at the least, 105 base pairs (such as *vestigial*$^{21-1.3}$) of the 5' *P* element sequences (FIGURE 4.1 iii, iv and v). The region between the 5' inverted-repeat and position 105 contains both the promoter and the transcriptional start site of the *P* element transcription unit (KARESS and RUBIN 1984). The above deletion analysis suggests that the presence of the *P* element promoter may be responsible, to some extent, for generating the more extreme *vestigial* phenotypes from *vestigial*^{21}.

The derivatives that display a strong *vestigial* phenotype can be subdivided into two groups, the strap-wing (FIGURE 4.1 iii and iv) and the classic *vestigial*-wing classes (FIGURE 4.1 v). Six strap-wing lines were isolated which fall into two distinct categories. Three alleles (*vestigial*$^{21-1.3}$, *vestigial*$^{21-161.1}$ and *vestigial*$^{21-213.1}$) are separate isolates but share an identical lesion which has one breakpoint in the 3' terminal repeat and another close to the *P* promoter (FIGURE 4.1 iv). The other strap alleles (*vestigial*$^{21-6.1}$, *vestigial*$^{21-74.2}$ and *vestigial*$^{21-213.4}$) have the smallest internal deletions found among the strong *vestigial* derivatives (FIGURE 4.1 iii). All three alleles are missing the site where polyadenylation occurs. The site of the *P* element transcript polyadenylation (LASKI *et al.* 1986) has been excised in all alleles that display the classic or strap-wing *vestigial* phenotypes. Conversly, the presence of the transcription termination signal within the *P* element, as in *vestigial*^{21}, only moderately effects *vestigial* expression and results in the cryptic phenotype.

Northern analysis of *P* element alleles of *vestigial*: To examine the effect of the presence of *P* element sequences upon *vestigial* transcription, Northern analysis was carried out on several *P* element alleles of *vestigial* (FIGURE 4.2). The Oregon R strain, the wild type stock from which *vestigial*^{21} was derived, produces a 3.8 kb transcript (WILLIAMS *et al.* 1990a). The four *P* element containing alleles, *vestigial*$^{21-1.3}$, *vestigial*$^{21-60.1}$, *vestigial*$^{21--6.1}$ and *vestigial*^{21}, each produce a transcript that appears larger than the wild type *vestigial* transcript. The *vestigial*^{21} line which bears the largest

FIGURE 4.2. Northern analysis of *P* element alleles of *vestigial*. Total RNA from wild type flies and flies bearing four different *P* element insert alleles, *vestigial*$^{21-1.3}$, *vestigial*$^{21-60.1}$, *vestigial*$^{21-6.1}$ and *vestigial*^{21}, was fractionated by electrophoresis, transferred to a nitrocellulose support and allowed to hybridize with an oligolabled DNA probe. The probe was generated from the *Eco*RI-*Hinc*II fragment of *vg* cDNA2 which corresponds to sequences from exons 1, 2 and 3 of the 3.8 kb *vestigial* mRNA. The Oregon R strain has a 3.8 kb transcript (WILLIAMS *et al.* 1990a). The four *P* element alleles each produce a transcript that appears larger than the wild type *vestigial* transcript. The *vestigial*^{21} line produces the largest message, the *vestigial*$^{21-1.3}$ and *vestigial*$^{21-60.1}$ mutants produce transcripts that appear smaller than those of the other *P* element alleles but are larger than wild type and the *vestigial*$^{21--6.1}$ allele appears to yield a band of intermediate length. Size markers are not included but the wild type message has been shown to be 3.8 kb (WILLIAMS *et al.* 1990a).


OrR
vg21-1.3
vg21-60.1
vg21-6.1
vg21
3.8 kb -

insert of the alleles tested, produces the largest message. The *vestigial*$^{21\text{-}1.3}$ and *vestigial*$^{21\text{-}60.1}$ mutants, which have the smallest *P* elements, produce transcripts that, although larger than the wild type transcript, appear smaller than those of the other *P* element alleles. The *vestigial*$^{21\text{-}6.1}$ allele appears to yield a band smaller than *vestigial*^{21} but larger than the *vestigial*$^{21\text{-}1.3}$ and *vestigial*$^{21\text{-}60.1}$ alleles. The presence of mRNAs larger than 3.8 kb is consistent with the existence of a *P* element-*vestigial* fusion message which may suggest that transcription initiated at the *P* element promoter continues into the *vestigial* transcription unit in some *P* element alleles. A possible structure of such a hybrid transcript is presented in FIGURE 4.3.

DISCUSSION

Although necessarily active in the germline, the *P* element promoter has been demonstrated to be active in somatic tissues (LASKI *et al.* 1986; O'KANE and GEHRING 1987). Transcription from the *P* element does not inevitably terminate at the *P* element polyadenylation signals but can continue into adjacent genomic sequences (LASKI *et al.* 1986). *P* element expression from a *Pc[ry]*^{2} transformant produced two messages, a 2.5 kb transcript terminating at the *P* element's polyadenylation site and a 3.0 kb transcript ending in the adjacent *rosy* sequences. Another, much larger band, which may represent another read-through product is apparent in the lane of a Northern containing *Pc[ry]*^{2} embryos mRNA (LASKI *et al.* 1986). The activity of the *P* element promoter in somatic tissues and the ability of transcription to carry on past the termination signals of the *P* element seem to be responsible for the characteristics of the *P* element alleles of *vestigial*.

It is possible that transcriptional interference of the *vestigial* promoter, which is located several hundred base pairs downstream of the *P* element insertion site (See CHAPTER 5), by the product of the *P* element promoter is responsible for the properties of *vestigial*^{21} and its derivatives (FIGURE 4.4). This notion correlates well with the mutant phenotypes of *vestigial*^{21} and its derivatives, as well as with the respective associated molecular lesions. This is not a new suggestion, as transcriptional interference was originally proposed to account for the phenotypes of the *vestigial P* element alleles. However, the original study proposed a mechanism by which *P* element activity would decrease the transcription of a message transcribed in the opposite direction (WILLIAMS *et al.* 1988a). The message originating within the second *vestigial* intron has since been demonstrated to be unimportant to *vestigial* function (WILLIAMS *et al.* 1990a). Sequence analysis of the extreme derivatives demonstrates that the presence of the *P* promoter

FIGURE 4.3. A possible *P* element fusion transcript at *vestigial.* A restriction map of the *vestigial* genomic region and an exon map of the 3.8 kb *vestigial* transcript is presented (adapted from WILLIAMS and BELL, 1988). The structure of a predicted *P* element-*vestigial* fusion transcript is presented below. The exons are aligned with the genomic sequences and one another.

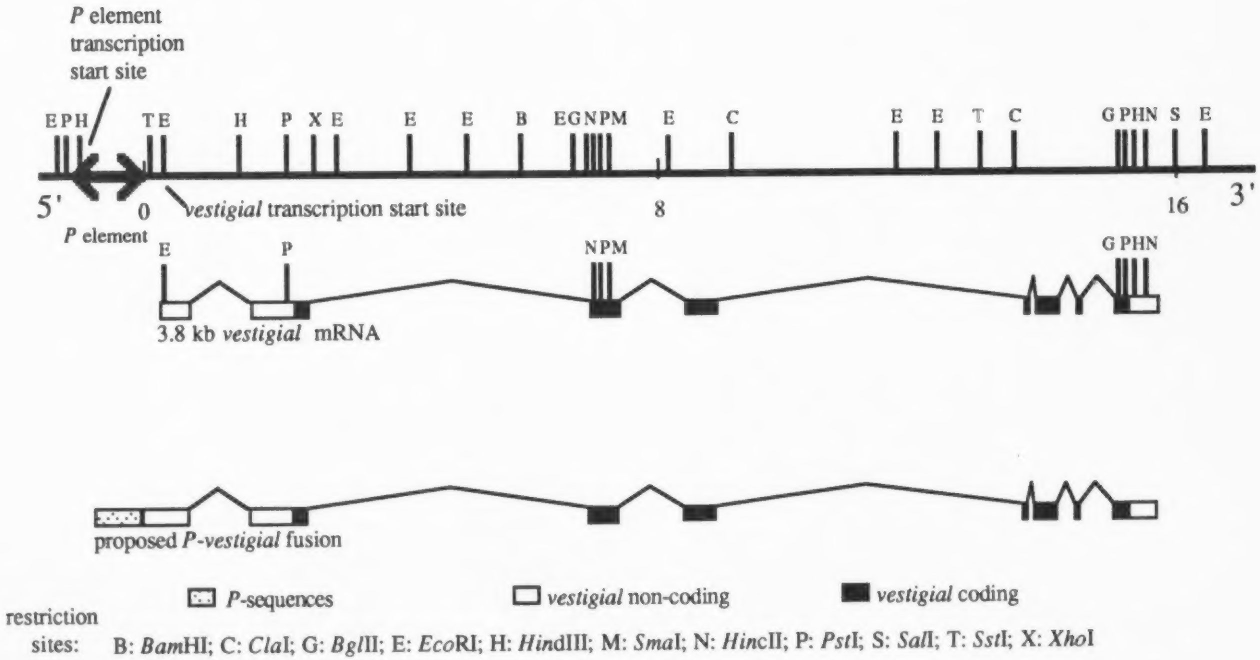
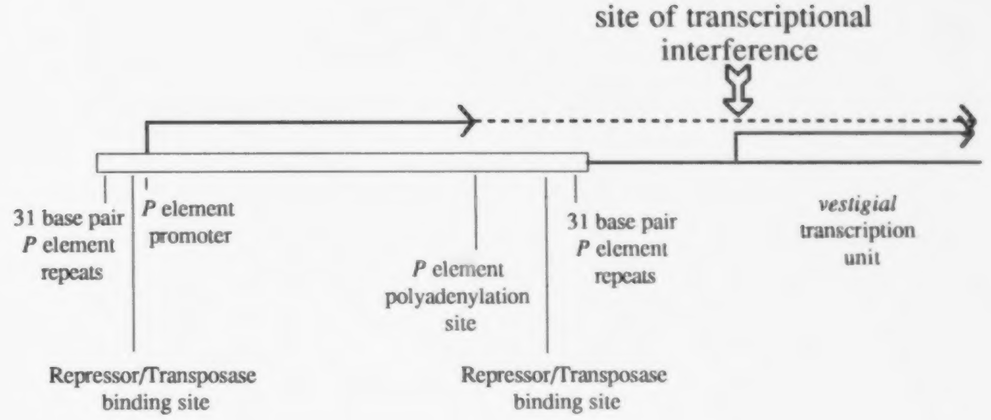
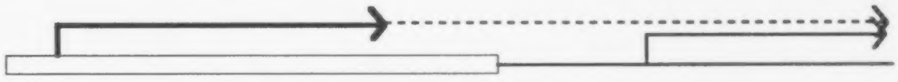

P element transcription start site
E P H
T E
H
P
X
E
E
E
B
E G N P M
E
C
E
E
T
C
G P H N
S
E
5'
0
vestigial transcription start site
8
16
3'
P element
E
P
N P M
G P H N
3.8 kb vestigial mRNA
proposed P-vestigial fusion
P-sequences
vestigial non-coding
vestigial coding
restriction sites:
B: BamHI; C: ClaI; G: BglII; E: EcoRI; H: HindIII; M: SmaI; N: HincII; P: PstI; S: SalI; T: SstI; X: XhoI

FIGURE 4.4. A transcription interference model of *vestigial*^{21} and its derivatives. Transcription initiated at the *P* element promoter may interfere with *vestigial* transcription. If the *P* transcript is usually terminated within the *P* element then the influence upon *vestigial* transcription is minimal (ie. *vestigial*^{21}). In revertant alleles of *vestigial*^{21} the *P* promoter is removed and the *vestigial* transcription unit is again normal. The absence of the 3' region of the *P* element may result in extensive read through of the *P* element message and decreased *vestigial* expression. Cytotype repression may be accomplished by binding of the repressor to the *P* element promoter to reduce the initiation of transcription. In addition, read through may be reduced by either the presence of an active polyadenylation signal or the repressor protein interacting with the binding site inside the 3' end or both. The *P* element landmarks are indicated below the diagram.

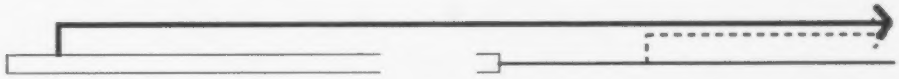

site of transcriptional interference
31 base pair P element repeats
P element promoter
P element polyadenylation site
31 base pair P element repeats
vestigial transcription unit
Repressor/Transposase binding site
Repressor/Transposase binding site

i



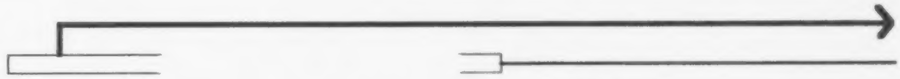
vestigial-21

ii



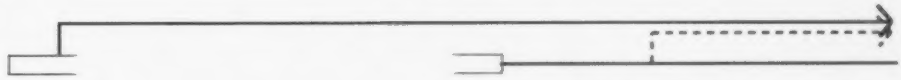
strap wing derivatives (small internal deletion) ie. $vg^{21-6.1}$, $vg^{21-74.2}$, $vg^{21-213.4}$

iii



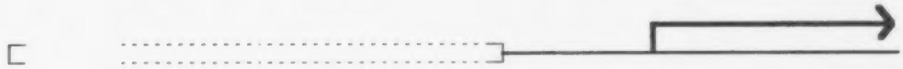
classic *vestigial* derivatives ie. $vg^{21-33.1}$, $vg^{21-60.1}$, $vg^{21-83.1}$

iv



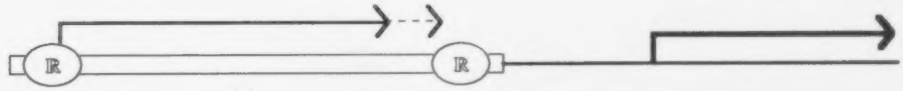
strap wing derivatives (large internal deletion) ie. $vg^{21-1.3}$, $vg^{21-161.1}$, $vg^{21-213.1}$

v



revertants ie. vg^{21-IR2}, vg^{21-IR7}, vg^{21-IR8}, $vg^{21-IIIR3}$

vi



cytotype repression ie. vg^{21} (P cytotype) without the 3' *P* element sequences, including the *P* element polyadenylation signal, may lead to reduced *vestigial* function. In the revertant alleles, no *P* element promoter is present and the *vestigial* activity approaches wild type levels. Transcription initiated from the *P* element promoter may interfere with normal *vestigial* expression. In *vestigial*^{21}, transcription may be initiated at the *P* promoter and terminated predominantly at the polyadenylation site. Thus transcription can then also be reinstated downstream at the *vestigial* transcription start site. The Northern analysis does not clearly demonstrate the presence of the normal *vestigial* transcript in *vestigial*^{21} and further experiments are required. This series of events should act to influence *vestigial* expression only slightly and result in complete wing formation in the *vestigial*^{21} homozygote and a 'scalloped wing' in the hemizygote. In the absence of *P* transcript termination within the *P* element, *vestigial* expression is greatly reduced. In summary, transcription initiated at the *P* element promoter may continue through the *vestigial* start site and cause a reduction in *vestigial* function. The presence of the polyadenylation site, as in *vestigial*^{21}, may act to terminate *P* element transcription and thus cause only a slight decrease in *vestigial* function. Deletion of the *P* promoter returns *vestigial* expression to normal levels even if several hundred base pairs of *P* element sequences remain at the original insertion site of *vestigial*^{21}.

The interference model proposed herein is dependent upon transcription from the *P* promoter to produce a product that reduces the production of a normal *vestigial* transcript. The presence of the *P* element-*vestigial* fusion would satisfy the above conditions, if the fusion fails to compensate for decreased expression of the nornal *vestigial* transcript. The original *vestigial*^{21} allele retains several hundred base pairs of both the 5' and 3' *P* element ends (FIGURE 4.4 i) and *vestigial* expression is only slightly lower than wild type. Of the more extreme *vestigial* alleles derived from *vestigial*^{21}, the homozygous strap-wing alleles comprise a minor phenotypic class (STAVELEY *et al.* 1995). Three strap-wing alleles (*vestigial*$^{21\text{-}6.1}$, *vestigial*$^{21\text{-}74.2}$ and *vestigial*$^{21\text{-}213.4}$) have small deletions of the 3' *P* element and do not retain the polyadenylation site (FIGURE 4.4 ii). Loss of this sequence appears to greatly reduce *vestigial* expression. However, *P* alleles that retain sequences between position 294, the 5' breakpoint of *vestigial*$^{21\text{-}33.1}$ ("classic" *vestigial* phenotype) and position 2697, the 5' breakpoint of *vestigial*$^{21\text{-}213.4}$, display the strap-wing phenotype (FIGURE 4.4 iii). This may indicate that failure to terminate transcription at the 3' end of the *P* element can be partly compensated by the sequences between these sites. The phenotypes of the three strap-wing alleles (*vestigial*$^{21\text{-}1.3}$, *vestigial*$^{21\text{-}161.1}$ and *vestigial*$^{21\text{-}213.1}$) which share an even larger internal deletion and a breakpoint at nucleotide 105 may reflect a slight decrease in the activity of the *P* promoter (FIGURE 4.4 iv). These mutants appear to delete sequences between positions 105 and 110 necessary for the maximum interference of the *vestigial* promoter and, possibly, for normal levels of *P* element transcription. Deletions of the *P* element promoter (FIGURE 4.4 v) appear to completely abolish *P* element transcription and allow normal *vestigial* expression. This model accounts for the phenotype of most derivatives of *vestigial*^{21}.

The cytotype dependent behavior of *vestigial*^{21} and some derivatives has been demonstrated (WILLIAMS *et al.* 1988b). The *vestigial*^{21} allele is somewhat sensitive to the *P* cytotype and displays complete wings as a heterozygote over a null allele, instead of the strapped to nicked wing phenotype seen in the *M* cytotype. The most sensitive allele (*vestigial*$^{21\text{-}3}$) has been utilized to report the cytotype activity of *P* element repressor activity (GLOOR et al,. 1993). The suppression of the *vestigial*$^{21\text{-}3}$ phenotype in the *P* cytotype has been proposed to be due to an increase in *vestigial* activity directed by the 66 kDa somatic repressor protein (RIO 1989). RIO suggested that the 66 kDa protein might activate gene expression from certain *P* element insertions, including the 66 kDa encoding elements, to provide a positive feed-back mechanism for the synthesis of the repressor protein.

The model of transcriptional interference of the normal *vestigial* transcript by the *P* element-*vestigial* fusion can be readily extended to account for the cytotype sensitivity of some *vestigial*^{21} derivatives (FIGURE 4.4 vi). When no *P* repressor molecules are present, the *P* promoter initiates transcription of the fusion transcript. The 5' binding site of the repressor molecule overlaps the *P* promoter and transcription of the fusion message should be repressed in the presence of the repressor molecule (KAUFMAN *et al.* 1989; KAUFMAN and RIO 1992). If *P* element repressor molecules bind at the *P* element promoter, transcription of the *P* element-*vestigial* fusion transcript may not be initiated. This should relieve transcriptional interference of the normal *vestigial* promoter and, as a result, *vestigial* expression could approach wild type in a *P* background. This is similar to the effect of generating *P* element revertant alleles that do not retain the *P* element promoter. Deletion alleles missing the 3' sequences of the *P* element, including the polyadenylation site and the 3' transposase binding site, do not appear to be sensitive to the presence of *P* elements (WILLIAMS *et al.* 1988a; b). The combination of reduced transcription initiation at the *P* promoter, the presence of the polyadenylation site and the binding of the repressor to the 3' binding site (KAUFMAN *et al.* 1989) could effectively reduce transcription within the *P* element to where the *vestigial* promoter escapes all interference.

An examination of the phenotypes associated with the cytotype-sensitive *vestigial* alleles and a molecular characterization of their respective *P* element sequences may provide clues to the differences in phenotype observed. The original *P* element allele, *vestigial*^{21}, which possesses an internally deleted 0.7 kb *P* element acts as only a slight hypomorph in the *M* cytotype and is totally suppressed in the *P* cytotype (WILLIAMS *et al.* 1988b). The most sensitive allele, *vestigial*$^{21\text{-}3}$, has an extreme phenotype in the absence of *P* elements but can be completely suppressed in the presence of *P* element activity. The latter allele has a much larger but incomplete *P* element and the difference in the phenotypes probably resides within the additional sequences. Based upon sequences present, this element should be able to produce a type I repressor (GLOOR *et al.* 1993). However, due to the sensitivity of *vestigial*$^{21\text{-}3}$ to the presence of *P* element repressors and the mechanism proposed to explain this phenomenon, it is unlikely that *vestigial*$^{21\text{-}3}$ produces much if any repressor. Another *P* element suppressible allele, *vestigial*$^{21\text{-}7Rev}$, resulted from the recruitment of an incomplete 1.1 kb *P* element deleted for sequences between positions 211 and 1968. Similar to *vestigial*^{21} , this allele has both the *P* element promoter and the polyadenylation site and displays a sensitivity to *P* elements, but is only a slight hypomorph in the *M* background. At least some of the sequences which allow *vestigial*$^{21\text{-}3}$ to produce an extreme phenotype in the *M* cytotype are located between the 5' breakpoint of *vestigial*^{21} (position 380) and the 3' breakpoint of *vestigial*$^{21\text{-}7Rev}$ (position 1968). This region includes the 3' region of ORF 0, ORF 1 and all of ORF 2. Contained within these sequences are several *cis*-acting signals, including the splice donor sites of ORF 0, ORF 1 and ORF 2, which may be responsible for the *vestigial* phenotype of *vestigial*$^{21\text{-}3}$. A detailed analysis of the transcription products of this allele may be required to distinguish between this and other possible causes.

The suppression of several *singed* alleles in a *P* cytotype appears to be analogous to the cytotype-dependent behaviour of the suppressible *P* element alleles of *vestigial*.. A number of alleles of *singed* containing defective *P* element inserts (ROBERTSON *et al.* 1988) were tested for a cytotype-dependent bristle morphology (ROBERTSON and ENGELS 1989). Examination of the sons of *singed* mutant males and compound X females, of both *P* and *M* stocks, demonstrated that three alleles (*sn*^{50e}, *sn*^{91e} and *sn*^{103e}) displayed extreme *singed* bristles in the *M* cytotype and a moderate bristle phenotype in the *P* cytotype. Interference of *singed* transcription by *P* element activity may be responsible for the phenotypes observed in both *P* and *M* backgrounds. The *P* cytotype can also influence the expression of the white gene inserted within a *P* element. An *in vitro* modified *P* element, *P*[w^{dl}9.3] (19DE), contains the complete *white* gene along with a tandem direct duplication of the 5' regulatory region, containing the sequences required for *white-zeste* interaction, and the first exon (COEN 1990). The near wild type expression of the transgene in the *M* cytotype is modified to a brown eye colour (paler in the posterior of the eye) in a *P* cytotype. The reduction of the pigment level was not different in reciprocal crosses, but only dependent upon the presence of *P* element expression. The *P* repression is strongly contingent upon the insertion site of this transgene. The transgene was mobilized and only 1 of 77 mobilization events (one at 36AB) resulted in another *P* suppressible insertion. This is apparently a position-dependent phenomenon. It is possible that the insert at 19DE and 36AB places the *P* element promoter under the control of *cis*-acting elements which interferes with the expression of the *white* product. This question was addressed by Northern blot analysis which revealed approximately equal expression in *P* and *M* strains of P[w^{dl}9.3] (19DE). Transcription initiated by the *P* element promoter driving *white* to near wild type levels in the *M* cytotype and less so in the *P* cytotype may not have been uncovered if the difference is tissue specific or too subtle to be detected by Northern analysis.

The proposed mechanism responsible for the phenotypes of *vestigial*^{21} and its derivatives suggest that a transposable element can exert control over the transcription of *vestigial*. The derivatives of this allele represent several alternative secondary outcomes. The revertant alleles are relieved of the negative effects of the invading transcription and the extreme alleles have been placed under a greater influence of the transposon's promoter. The cytotype dependent alleles are conditionally either relatively free from the effects of the transposon's promoter or under its tight control. Although a product of artificial selection, the presence or absence of wings in the *vestigial*$^{21-3}$ allele is greatly dependent upon the presence of a transacting factor, the *P* repressor. The expropriation of a coding domain by an adjacent transposable element's control region may be a common and unstable source of genetic variation.

Studies of the influence that *P* elements exert upon the expression of the *lacZ* gene fused to the *P* promoter have provided a model for the control of *P* element gene fusions by the products of the *P* element. Control of *P*-*lacZ* enhancer trap expression of ß-galactosidase has demonstrated that the *P* element-reporter gene fusion is susceptible to the effects of *P* regulatory products in the soma (LEMAITRE and COEN 1991). All of the

P-lacZ insertions analyzed revealed a reduction, but not the complete absence, of *lacZ* expression in a *P* background. As a control, the *hsp70-lacZ* fusion enhancer trap did not reveal a sensitivity to cytotype. *P-lacZ* expression in the germline was investigated through the study of enhancer trap insertions that lead to recessive female sterility, all of which display ß-galactosidase expression in the ovaries and testes (LEMAITRE *et al.* 1993). Germline expression of *P-lacZ* did not differ in the germline of the progeny of reciprocal crosses when crossed to *M* cytotype strains. However, crosses to *P* strains revealed high levels of ß-galactosidase activity in the germline of the offspring of *P* males and *M* females . This activity is, however, repressed in the progeny of the reciprocal cross. This activity is apparently directly controlled by the *P* repressor and does not involve secondary interference such as proposed here.

The existence of a transposable element which can act as a mobile promoter bears some examination. The consequences of a transposon of this nature being scattered throughout the genome could result in a number of genes being placed under the influence of the same set of transcriptional control mechanisms. This could certainly make a population respond quite differently and immediately to the presence or absence of *P* element proteins. For example, introduction of a *P* element into a population that is cytotype-dependent for the development of wings could greatly influence the dynamics of a population and its response to selection. Analysis of a number of genes controlled by the *P* promoter, under various selection schemes, could be developed as a model system by which the development of complex transcriptional mechanisms, such as the origin of a cascade of transcription factors, could be examined.

LITERATURE CITED

COEN, D., 1990 *P* element regulatory products enhance *zeste*^{1} repression of a *P*[*white*^{duplicated}] transgene in *Drosophila melanogaster*. Genetics **126**: 949-960.

ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 437-484 in *Mobile DNA*., edited by D.E. BERG and M.M. HOWE. American Society for Microbiology, Washington DC.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS, W. K. BENZ, H. M. ROBERTSON and W. R. ENGELS, 1993 Type I repressors of *P* element mobility. Genetics **135**: 81-95.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131**: 917-927.

HESLIP, T. R., and R. B. HODGETTS, 1994 Targeted transposition at the *vestigial* locus of *Drosophila melanogaster*. Genetics **138**: 1127-1135.

KARESS, P. D., and G. M. RUBIN, 1984 Analysis of *P* transposable element functions in *Drosophila*. Cell **38**: 135-146.

KAUFMAN, R. E., R. F. DOLL and D. C. RIO, 1989 Drosophila *P* element transposase recognizes internal *P* element DNA sequences. Cell **59**: 359-371.

KAUFMAN, R. E., and D. C. RIO, 1992 *P* element transposition *in vitro* proceeds by a cut and paste mechanism and uses GTP as a cofactor. Cell **69**: 27-39.

LASKI, F. A., D. C. RIO and G. M. RUBIN, 1986 Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44**: 7-19.

LEMAITRE, B., and D. COEN, 1991. *P* regulatory products repress in vivo the *P* promoter activity in *P-lacZ* fusion genes. Proc. Natl. Acad. Sci. USA **88**: 4419-4423.

LEMAITRE, B., S. Ronsseray and D. COEN, 1993 Maternal repression of the *P* promoter in the germline of *Drosophila melanogaster*: A model for the *P* cytotype. Genetics **135**: 149-160.

LASKO, P. F., and M. L. PARDUE, 1988 Studies of the genetic organization of the *vestigal* microregion of *Drosophila melanogaster*. Genetics **120**: 495-502.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

MISRA, S., and D. C. RIO, 1990 Cytotype control of Drosophila *P* element transposition: the 66kD protein is a repressor of transposase activity. Cell **62**: 269-384.

MISRA, S., R. M. BURATOWSKI, T. OHKAWA and D. C. RIO, 1993 Cytotype control of *Drosophila melanogaster P* element transposition: genomic position determined maternal repression. Genetics **135**: 785-800.

MULLINS. M. C., D. C. RIO and G. M. RUBIN, 1989 Cis-acting DNA sequence requirements for *P* element transposition. Genes Dev. **3**: 729-738.

NASH, D., and J. B. BELL, 1968 Larval age and the pattern of DNA synthesis in polytene chromosomes. Can. J. Genet. Cytol **10**: 82-92.

O'HARE, K., and G. M. RUBIN, 1983 Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**: 25-35.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84**: 9123-9127.

O'KEEFE. S., M. SCHOULS and R. HODGETTS, 1995 Epidermal cell-specific quantification of *dopa decarboxylase* mRNA in Drosophila by competitive RT-PCR: an effect of Broad-complex mutants. Develop. Genet. **16**: 77-84.

RASMUSSON, K. E., J. D. RAYMOND and M. J. SIMMONS, 1993 Repression of hybrid dysgenesis in *Drosophila melanogaster* by naturally occurring *P* elements. Genetics **133**: 605-622.

RIO., D. C., 1990 Molecular mechanisms regulating Drosophila *P* element transposition. Ann. Rev. Genet. **24**: 543-578.

RIO, D. C., 1991 Regulation of Drosophila *P* element transposition. Trend. Genet. **7**: 282-287.

ROBERTSON, H. M. and W. R. ENGELS, 1989 Modified *P* elements that mimic the *P* cytotype in *Drosophila melanogaster*. Genetics **123**: 815-824.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ and W. R. ENGELS, 1988 A stable genetic source of *P*-element transposase in *Drosophila melanogaster*. Genetics **118**: 461-470.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular cloning: A laboratory manual*. Cold Spring Harbor Press, Cold Spring Harbor, New York.

SIMMONS, M. J., and L. M. BUCHOLZ, 1985 Transposase titration in *Drosophila melanogaster*: a model of cytotype in the *P-M* system of hybrid dysgenesis. Proc. Natl. Acad. Sci. USA **82**: 8119-8123.

STAVELEY, B. E., R. B. HODGETTS, S. L. O'KEEFE and J. B. BELL, 1994 Targeting of an enhancer trap to *vestigial*. Develop. Biology **165**: 290-293.

STAVELEY, B. E., T. R. HESLIP, R. B. HODGETTS and J. B. BELL, 1995 Protected *P* element termini suggest a role for IRBP in transposase-induced gap repair in *Drosophila melanogaster*. Genetics **139**: 1321-1329.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**:1355-1363.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988a Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988b Suppressible *P*-element alleles of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **212**: 370-374.

WILLIAMS, J. A., A. L. AKTIN and J. B. BELL, 1990a The functional organization of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **221**: 8-16.

WILLIAMS, J. A., I. M. SCOTT, A. L. AKTIN, B. J. BROOK, M. A. RUSSELL and J. B. BELL, 1990b Genetic and molecular analysis of vg^U and vg^W: Two dominant *vg* alleles associated with gene fusions in Drosophila. Genetics **125**: 833-844.

WILLIAMS, J. A., S. W. PADDOCK, K. VORWERK and S. B. CARROLL, 1994 Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature **368**: 299-305.

Chapter 5

Reexamination of *vestigial* expression

The *vestigial* gene has long been known to be important in wing development (MORGAN 1911). The Drosophila thorax has appendages which consist of the dorsal wings and halteres and three ventral pairs of legs. During the development of the imaginal discs, which give rise to the appendages, three distinct boundaries are formed during a step wise series of compartmentalization events. Three axes are established in the following order: first the anterior/posterior, then the dorsal/ventral and finally the proximal/distal axis (WILLIAMS and CARROLL 1993; COHEN 1993). One class of genes involved in both dorsal and ventral discs to direct the formation of distal structures includes *decapentaplegic* (*dpp*), *wingless* (*wg*) and *aristaless* (*al*). Other genes are required for the patterning of the proximal/distal axis in either the leg or wing discs but not both. *Distal-less* (*Dll*) is involved in the establishment of the proximal/distal compartments in the leg discs but not in the dorsal wing and haltere discs (COHEN and JURGENS 1989). In contrast, the *vestigial* (*vg*), *scalloped* (*sd*) and *apterous* (*ap*) genes are not required for leg development but are vital for the formation of the proximal/distal axis in the wing and the loss of function of any of these genes results in, at the very least, the loss of distal structures in the wing (WILLIAMS *et al.* 1993; CAMPBELL *et al.* 1992; COHEN *et al.* 1992). Absence of *vestigial* function results in extensive cell death (FRISTROM 1968; 1969) and was proposed to be due to abnormal positional information in the wing disc (BOWNES and ROBERTS 1981). Recently, the *vestigial* gene has been demonstrated to function in the establishment of positional cues in wing formation (WILLIAMS *et al.* 1993; 1994). In the wing disc, early expression of *wingless* and *apterous* acts to produce the dorsal/ventral boundary. The *apterous* protein is restricted to the dorsal portion of the imaginal wing disc as an essential step in the specification of the dorsal fate. As a result, interactions between dorsal and ventral cells result in expression of *vestigial* in a narrow stripe of cells at the dorsal/ventral border (WILLIAMS *et al.* 1991; 1994). The *vestigial* gene is clearly central to the development of the Drosophila wing.

The *vestigial* gene was cloned via *P* element transposon tagging (WILLIAMS and BELL 1988). The mutational lesions associated with a number of *vestigial* mutants were mapped to a 19 kb region within polytene section 49D-E. A 3.8 kb transcript of low abundance, identified by hybridization to several genomic probes from the *vestigial* region, and several *vestigial* cDNAs were isolated from an imaginal disc library (WILLIAMS *et al.* 1990). The 3.8 kb mRNA was found to be present in embryos, early larvae and pupae. A sequence representing the *vestigial* transcription unit and flanking regions was reported, which predicted a novel nuclear protein (WILLIAMS *et al.* 1991). However, no functional similarities to known proteins were noted. The *vestigial* message shows a complex pattern of expression, but most notably is expressed in the early precursors of the wing and haltere discs. Antibodies to recombinant *vestigial* protein reveal that expression in the late third instar imaginal disc occurs along a thin line at the presumptive wing margin (WILLIAMS *et al.* 1991; 1993). An enhancer element has been identified in the second intron which is apparently responsible for expression along the dorsal/ventral boundary of the wing margin (WILLIAMS *et al.* 1994) and, based upon the identification of another highly conserved intronic region, a second transcriptional control element may be present in the fourth intron (WILLIAMS *et al.* 1991). The nature of the *vestigial* gene product and its function in the process of determining the wing and wing margin is presently unclear.

Alternative splicing, the generation of different transcripts from a single pre-mRNA, is significant in the control gene expression of eukaryotes (MCKEOWN 1992; HODGES and BERNSTEIN 1994). Regulation of alternative splicing is often a reflection of tissue-specific or developmentally-specific control strategies. The control exerted over the expression of *vestigial* appears to be complex and includes the transcriptional control elements described above. Although a 3.8 kb transcript was identified in the previous report of *vestigial* transcription products, Northern analysis revealed a number of other transcripts which hybridize to at least one *vestigial* probe used in this study. These transcripts were attributed to cross hybridization of non-*vestigial* mRNAs which share regions of homology with the probe used. To address the alternate hypothesis, the possibility that *vestigial* may produce more than one message, the transcription of *vestigial* has been reexamined. Products of reverse transcription have been isolated which suggest that the *vestigial* pre-mRNA undergoes alternative splicing. In addition, the transcriptional start site has been mapped and several additional features of the transcript have been identified including a potential alternative open reading frame and several structural motifs that may be important in transcription, splicing and translation. A model of the influence of alternative splicing upon *vestigial* expression is presented.

MATERIALS AND METHODS

Drosophila melanogaster stock: The Canton S stock, a standard laboratory wild type strain, was grown on synthetic sugar/yeast/agar medium (NASH and BELL 1968) at room temperature (~22°C).

Recombinant DNA technology: DNA manipulations were performed according to standard methods (SAMBROOK *et al.* 1992). The restriction map of the 45 kb genomic region including *vestigial* (WILLIAMS and BELL. 1988) and the sequence of the *vestigial* transcription unit and flanking genomic DNA (WILLIAMS *et al.* 1991) have been previously reported. Six *vestigial* cDNAs were isolated (WILLIAMS *et al.* 1990) from a third instar larval imaginal disc cDNA library provided Dr. G. RUBIN. Due to the extinction of the original subclones, the insert of cDNA6 was re-subcloned into pT7T3-19U to facilitate sequencing. Some of the amplified products were cloned into pGEM-T vector (Stratagene).

Primers: Single-stranded deoxynucleotide primers were used to initiate reverse transcription in the synthesis of first strand cDNA from the mRNA templates. The primer is complementary to the 5' region of the second exon (TTCCCTTGGCGTTGGAACG) in the RACE protocol. In the RACE procedure, a nested second exon primer, upstream from the sequence described above and next to the splice donor site (TCCGACGAAGAATTGCCAA), the 5' RACE anchor primer (CUACUACUACUAGGCCACGCGTCGACTAGTACGGGIIGGGIIG) and the 5' RACE amplification primer (TACTAGGCCACGCGTCGACTAGTA) were used.

RNA isolation: Crude total RNA extracts were prepared as described by O'KEEFE *et al.* (1995). Ten white pupae were added to the RNA isolation buffer composed of 0.1 M Tris, 0.1 M NaCl, 20 mM EDTA and 1% SDS in a sterile, baked RNase-free glass homogenizer. After one quick plunge of the pestle, an equal volume of phenol equilibrated with sterile DEPC-treated distilled water was added to the extraction mixture. Homogenization of the tissue was continued in the phenol/buffer mixture for a total of ten plunges. The extraction mixture was then transferred to microfuge tubes and the phases separated by a 30 second 12000 rpm microfuge centrifugation. The aqueous phase was removed, carefully avoiding contamination of material from the phenol/water interface. The aqueous phase was quickly extracted three times with equal volumes of diethyl ether. To the extracted sample, 2.5 volumes of cold ethanol was added and precipitation was carried out over 30 minutes at room temperature. After a 5 minute microcentrifugation, aspiration of the remaining liquid and quick drying under low pressure, the pellet was resuspended in 20 ul of DEPC-treated water (2 ul per animal) and stored at -80C until further manipulation.

First strand cDNA synthesis: The synthesis of a reverse transcription product is essential to both the RT/PCR and RACE protocols. Reverse transcription (see O'Keefe *et al.* 1995), dC tailing of the cDNA and amplification were performed as according to manufacturer's instruction (Gibco BRL). To 1 ul of the mRNA preparation 1 ul of primer (100 ng/ul) and 12 ul of DEPC-treated distilled water was added. The mRNA was allowed to denature at 70 C for 5 minutes then chilled on ice for 1 minute. To this 1 ul of RNasin (Promega) was added. According to 5' RACE System instructions, 2.5 ul of a 10 X reaction buffer, 3 ul of 25 mM $MgCl_2$ 1 ul of 10 mM dNTP mix and 2.5 ul of 0.1 M DTT were added, gently mixed, collected by brief centrifugation and pre-incubated for 2 minutes at 42 C. To this 1 ul of Superscript II RT (Gibco BRL) was added, gently mixed and incubated for 30 minutes at 42 C. To terminate this reaction and remove the RNA, this was then denatured at 95 C for 10 minutes and placed at -80 C.

dC tailing of first strand cDNA: To mediate annealing of the anchor primer, a stretch of cytosine residues are added to the 3' end of the cDNA produced above. The cDNA preparation was pre-incubated at 55 C for 1 to 2 minutes 1 ul of RNase H was added, incubated at 55 C. To 10 ul of the first strand cDNA preparation, 7.5 ul of DEPC-treated water, 2.5 ul of 10 X reaction buffer 1.5 ul of 25 mM $MgCl_2$ and 2.5 ul of 2 mM dCTP were mixed and collected by centrifugation. The reaction mix was denatured for 2 to 3 minutes at 94 C, chilled on ice for 1 minute and again collected by centrifugation. To this 1 ul of terminal deoxynucleotidyl transferase (Gibco BRL) was added and incubated for 10 minutes at 37 C. The enzyme was heat inactivated by heating to 70 C for 10 minutes before icing.

RACE: The primers for the following procedures are described above. The amplification reactions were performed in a total volume of 30 ul starting with 3 ul of the cDNA preparation and a final concentration of 50 mM Tris-Cl (pH 9.2) 1.5 mM $MgCl_2$, 0.01% dithiothreitol , 0.1 ug/ul bovine serum albumin (Boehringer Mannheim), 200 uM of dATP, dCTP, dGTP and dTTP, 0.3 ng/ul of primer and 1U of *Taq* DNA polymerase (Gibco BRL). The amplifications were produced utilizing a Stratagene Robocylcer 40 with a program of 1 cycle of 5 min at 95°C 1 min at 60°C and 3 min at 73°C, followed by 30 cycles of 1 min at 92°C 1 min at 60° C and 3 min at 73°C. Products of the amplification were fractionated by electrophoresis on 1 to 2% agarose gels in standard Tris-acetate-EDTA buffer.

Sequencing of DNA: Manual DNA sequencing was performed using a Sequenase II DNA kit (USB) and separation of the products by electrophoresis on a 6% polyacrylamide gel. The cDNA6 insert cloned into pT7T3-19U and PCR products cloned into pGEM-T were subjected to double stranded plasmid sequencing according to manufacturer's instructions. PCR products were directly sequenced as described elsewhere (STAVELEY *et al.* 1995).

RESULTS

Examination of the 5' end of *vestigial*: The genomic region at the 5' end of the *vestigial* gene has been examined extensively since it is the insertion site of the *P* element of *vestigial*^{21} allele (WILLIAMS *et al.* 1988a;b). Although transcription of *vestigial* has been previously examined, the transcriptional start site of *vestigial* has not been reported. The longest *vestigial* cDNA isolated, *vg* cDNA1, extends 678 base pairs 5' of the splice donor site of the first exon (WILLIAMS *et al.* 1990; 1991). The start site of *vestigial* had been tentatively placed upstream of this. There is, however, no sequence similar to the consensus Drosophila transcriptional start site in the characterized 247 base pairs preceding the 5' end of this cDNA.

The 5' extent of the *vestigial* transcript has been determined by the Rapid Amplification of cDNA Ends (RACE) procedure by utilizing nested primers complementary to part of the second exon. The size of the fragment (~280 base pairs), accounting for the length of the primers and the location of the nested second exon primer, places the start site at approximately 230 base pairs 5' of the first exon splice donor site. Just upstream of this region, which is likely a transcriptional start site of *vestigial* (FIGURE 5.1) is the sequence AAATAATCG (positions 772 to 780 of the *vestigial* transcription unit; WILLIAMS *et al.* 1991) which matches 6 of 9 positions of the -25/-30 consensus sequence (c/G/c)TATAAA(g/a)(g/c) of *Drosophila melanogaster* (ARKHIPOVA 1995). Just downstream of the "TATA" box is a potential promoter cap site at 32 bases (TTCACGTG) from the start of the "TATA" box, with six of eight residues matching the consensus (t/A/g)TCA(g/t)T(t/c)G (ARKHIPOVA 1995). This places the *vestigial* transcriptional start site 466 base pairs downstream of the most 5' sequences of *vg* cDNA1 (WILLIAMS *et al.* 1990). It is possible that alternate promoters start upstream to produce transcripts that carry the 5' sequences present in this isolate. However, since

ATTTTGTGGATTTAGCGAAAATATTTTCATTGAGTGCGCCGA

CCAAATGTTTGCTGCTCGTTTTCGGTTAAAAATAATCGCAGA

→

AGCCGCTGCGCCAAAATCTCAAGAA TTCACGTG GAATTCC

FIGURE 5.1. The 5' region of the *vestigial* transcription unit (nt-701 to nt-825 of published sequence; WILLIAMS *et al*., 1991). The solid underlined sequence is a 6/9 base match to Drosophila "TATA" box (ARKHIPOVA, 1995). The double underlined sequence represents a match 6/8 base to the consensus transcriptional start site (ARKHIPOVA, 1995). The arrow indicates the transcriptional start site. This places the major, if not sole, transcriptional start site 466 base pairs downstream of the 5' end of *vg* cDNA1 and 486 base pairs downstream of the *P* element insertion site of *vestigial*^{21}. The published sequence (WILLIAMS *et al*., 1991) assigned position 1 of the transcription unit to the first residue of an *Eco*RI site upstream of the above landmarks. The previous placement of *vg* cDNA1 on the genomic map had placed it only 20 base pairs downstream of the *P* element insertion site.

vg cDNA1 has an additional 177 base pairs of DNA at the 5' end which originate from the second exon, it is possible that the entire sequence upstream of the transcriptional start site is an artifact of the construction of the library. An alternative explanation is that a transcript of lesser abundance, not detected by Northern analysis, was copied prior to incorporation of exon 2 sequences into the cDNA. However, since this cDNA extends much further upstream than any of the other cDNA's isolated, it appears most likely that the upstream sequences are not part of the *vestigial* mRNA.

Exon 1-3 alternative splicing: Of the six *vestigial* cDNAs isolated in the original study (WILLIAMS *et al.* 1990), one (*vg* cDNA6) differed from the other isolates in that the restriction map revealed an apparent internal deletion of sequences from the 5' portion of the transcript. The region spanning the absent sequences was sequenced by using a first exon primer near the 5' splice site. The sequence (FIGURE 5.2) demonstrates that in this cDNA the first and third exons are precisely spliced together. This product, if it does not differ from the 3.8 kb transcript in other ways, would produce a message 3.0 kb in length. This length of message is consistent with a band revealed by the Northern analysis (WILLIAMS *et al.* 1990). The intensity of the 3.0 kb band suggests that its level of expression is similar to that of the 3.8 kb band. The isolation of only one cDNA missing internal exon sequences in the earlier study may reflect either a disc-specific expression profile or a bias in the construction and/or screening of the library.

An amino acid sequence for the *vestigial* protein has been deduced by conceptual translation, antibodies have been made to detect it (WILLIAMS *et al.* 1991) and the expression pattern has been demonstrated to coincide with key aspects of wing development (WILLIAMS *et al.* 1993). The second exon contains the location of the translational start site of the conceptual *vestigial* protein (WILLIAMS *et al.* 1991) where two possible start codons are separated by 24 nucleotides (TABLE 5.1). Examination of the DNA sequence reveals that the first methionine codon is preceded by the best match to the Drosophila translational start consensus (CAVENER 1987). Alternative start codons do exist in the upstream region of exon 1, however all have in frame termination codons before the exon 1/intron 1 boundary. The third exon also possesses a potential a start codon which, if translation were to initiate there, would produce a truncated 317 residue protein. However, the preceding sequence does not match the consensus. Only the sixth AUG matches the consensus but this would result in a small (~60 amino acid) protein. The absence of the second exon would likely reduce or eliminate the production of

AAAACATTCAAGTG/TTCAACCAGGATTTG
TTTTGTAAGTTCAC/AAGTTGGTCCTAAAC

first exon donor site / third exon acceptor site

FIGURE 5.2. The *vestigial* exon 1/exon 3 splice junction. The sequence of the 5' region of *vg* cDNA6 reveals a precise splice between exon 1 and exon 3. This was determined by double-strand sequencing the cDNA insert cloned into pT7T3-19U using a *vestigial* first exon primer.

	Position of AUG	Exon	-4 to-1 sequence	Match to consensus	length of protein
1	1801	2	GACA	2/4	453
2	1825	2	AGTT	1/4	445
3	2212	3	GGGA	1/4	316
4	2596	4	GCCG	0/4	188
5	2770	5	CCCG	1/4	130
6	2983	6	CAAC	4/4	59
7	3097	7	ATGG	1/4	21

TABLE 5.1. Properties of the seven methionine codons of Vg, the *vestigial* protein (WILLIAMS *et al.*, 1991). The nucleotide position of all seven potential start codons of Vg are listed. The exon in which the codon is located, the nucleotide sequence preceding the methionine codon, the fraction of these nucleotides that match the translational consensus sequence (C/A)AA(A/C) (CAVENER 1987) and the predicted length of protein (translation ending at nucleotide 3162) are presented.

proteins in this reading frame. So it is likely that the translational start site located in the second exon is essential for the expression of this protein.

***vestigial* conceptual protein-2:** Both the full length message and the exon 1-3 spliced message could, theoretically, have the potential to encode another substantial product from a different reading frame. The amino acid sequence of this product, *vestigial* conceptual protein-2 (Vg-2) is presented in FIGURE 5.3. It is derived from the +1 reading frame, when compared to Vg, and would start translation in the fourth exon. The sequence CATAAUG, which includes the first methionine codon of Vg-2, is a very good match (6 of 7 residues) to the Drosophila consensus start site (CAVENER 1987). The positions of the start codons, the preceding nucleotide sequence and potential protein length are noted in TABLE 5.2. The proposed protein would be rich in proline, threonine and arginine. Also, Vg-2 has a stretch of similarity to PSF, a novel polypyrimidine tract-binding (PTB) protein-associated splicing factor (PATTON *et al.* 1993) isolated from HeLa cells (FIGURE 5.4). This region lies between the RNA-binding consensus domains of PSF and is not associated with a known function. The codon usage is less consistent with known Drosophila codon use than the protein predicted for Vg. Until actual experiments provide evidence for the existence of this protein, Vg-2 must be considered as a conceptual protein only. It is also possible that a novel product, a hybrid of the two proposed proteins due to perhaps a change in translational register, is the true product of the *vestigial* gene detected by antibodies. Since the existence of the second *vestigial* protein (Vg-2) is only hypothetical, the major effect of splicing out the second exon is probably the inability to code for a functional Vg protein.

The evidence presented above indicates that at least one message other than the 3.8 kb message (WILLIAMS *et al.* 1990; 1991) is produced by the *vestigial* locus. Re-examination of the developmental Northern analysis suggests that at least some of the transcripts originally described as artifacts are, most likely, alternate transcriptional products. The size of a number of the bands is consistent with a change in message length predicted by skipping several of the exons. In fact, the existence of the other transcripts may be indicative of the complexity of mechanisms which control *vestigial* expression. Further analysis is required to determine if other bands revealed by Northern analysis are alternate forms of the *vestigial* transcript.

```
MNHPA PPVRT PRATP TRTAI RCTVA AERRA VHLPR VVRAR GAATE AARGQ
YPRIC QHWRR RWAAE AADWP AVARD RLNIY RPPAL CSPTT RATRP ARWTS
TFPAP STTTT RILKR AAARC RIEIS HRRSG TAITC TPYPR PHTTR LATCM
APRRT PATPP IRGCR MRPTM VPMPT QRTPM RPTPT PTTTT WPSTA VSSGC
PSSTP ATVPG CTTTS RRPTP WSTPA IPQWQ AWKRR WRKYR NRRRI FTGSK
RHLAM EIADV LRRVY KPCPG RSKDT PDRIQ CATLG RDFES ASZ
```

FIGURE 5.3. The amino acid sequence of *vestigial* conceptual protein-2 (Vg-2). The hypothetical protein has a translational start site in the fourth exon and can encode a 292 amino acid protein. Vg-2 has a high arginine (17%), threonine (15%) and proline (13%) composition.

	Position of AUG	Exon	-4 to-1 sequence	Match to consensus	length of protein
1	2414	4	CATA	3/4	292
2	2961	6	TTGT	0/4	143
3	2909	6	CCGC	1/4	127
4	2921	6	CACT	2/4	123
5	2930	6	TCCT	0/4	120
6	2951	6	GCCC	1/4	113
7	3176	8	GGCG	0/4	38

TABLE 5.2. Properties of the seven methionine codons of a second conceptual *vestigial* protein (Vg-2). The nucleotide position of all seven potential start codons of Vg-2 are listed. The exon in which the codon is located, the nucleotide sequence preceding the methionine codon, the fraction of these nucleotides that match the translational consensus sequence (C/A)AA(A/C) (CAVENER, 1987) and the predicted length of protein (translation ending at nucleotide 3292) are presented.

```
PSF - ATPPTSGAPPGSGPGPTPTPPPAVTSAPPG
      a:::t:::pp :g : :pt  p  t:::p
Vg2 - APRRTPATPPIRGCRMRPTMVPMPTQRTPM

PSF - APPPTPPSSGVPTTPPQAGGPPPPPAAVPG
      :p:ptp:::: p:t: ::g p:::pa:vpg
Vg2 - RPTPTPTTTTWPSTAVSSGCPSSTPATVPG
```

FIGURE 5.4. A comparison of a 60 amino acid region of Vg-2 to PSF. The polypyrimidine tract-binding protein-associated splicing factor (PSF) is a human protein isolated from HeLa cells (PATTON et al., 1993). The region of homology is not yet associated with a function in the human protein. : represents conservative amino acid substitutions. The homology was discovered through utilizing the National Center for Biotechnolgy Information (NCBI) Blast electronic mail server.

DISCUSSION

In a previous study, developmental Northern analysis of *vestigial* expression revealed a series of bands which hybridize to two *vestigial* probes (WILLIAMS *et al.* 1990). These probes were a *Hin*cII-*Bgl*II DNA fragment and an antisense riboprobe, generated from a slightly smaller *Sma*I-*Bgl*II DNA fragment. Both fragments are subclones of *vg* cDNA1 and are specific for exons 3 to 8 of the *vestigial* cDNA. The 3.8 kb band, the only transcript detected by a DNA probe generated from an *Eco*RI-*Hin*cII of *vg* cDNA1 specific for exons 1 to 3, is apparent during embryonic, early larval and pupal stages but is only a minor component of the total mRNA that hybridizes to the probe. At least seven additional transcripts present at various stages of development hybridize to the *vestigial* riboprobe made to the *Sma*I-*Bgl*II DNA fragment described above. Inspection of the patterns of hybridization reveals two transcripts of approximately 3.0 kb and 2.5 kb in length which are expressed at the same time as the 3.8 kb message. These transcripts are very weakly expressed or undetectable during the second and third larval instar stages and in the adult. Also, a 2.0 kb band is very prominent during the embryonic and pupal stages. The smallest bands are either weakly expressed or absent in the embryo but are very apparent in the first, second and third instar larvae. The adult stage only displays one predominant band. If some or all of these bands represent legitimate *vestigial* transcripts, the expression patterns can be described as roughly reciprocal. The longest mRNAs, including the message identified as the complete 3.8 kb message, are present in embryos, young larvae and pupae. In contrast, the shorter transcripts are dominant during second and third larval instar stages.

The screening of an imaginal disc library resulted in the isolation of six independent *vg* cDNAs (WILLIAMS *et al.* 1990). The first cDNA (*vg* cDNA1) is the longest and extends further 5' than any of the others. The second, third and fourth *vestigial* cDNA's appear identical. The fifth cDNA apparently represents a prematurely truncated message and does not continue through to the 3' region of the transcription unit. The sixth *vg* cDNA is unusual and has homology to the 5' and 3' exons but is missing sequences which correspond to the second exon. Analysis of the 3' region revealed two overlapping polyadenylation sites (nt 3772 and nt 3777) flanked by two others (nt 3756 and nt 3808). The genomic location of the exons correlates well with the lesions associated with a number of the classical *vestigial* alleles.

Alternative splicing is an important means to regulate gene expression (for reviews see MCKEOWN 1992; HODGES and BERNSTEIN 1994). This process can be used to turn on and off gene functions and to generate alternate forms of a gene product having different functions. Alternative splicing may be manifested in several forms which include the decision to splice or not, the use of different promoters, alternative splice donor and/or acceptor sites, mutually exclusive exon use and single or multiple exon skipping. For example, the *Sex-lethal* (*Sxl*) gene of Drosophila produces a transcript which may or may not possess a specific exon based upon interpretation of the ratio of sex-chromosomes to autosomes. This exon bears an in-frame stop codon and inclusion of the exon results in a truncated protein which does not bind to its target sequence (BELL *et al.* 1988). Exon-skipping also occurs in the production of ecdysone receptor isoforms EcR-B1 and EcR-B2, wherein the latter arises when the second exon present in the former is spliced out (TALBOT *et al.* 1993). As a result, the B2 isoform does not possess the site from which the B1 isoform starts translation but employs another start codon unused by the longer mRNA. This results in two similar proteins that differ at their N terminal ends. The expression of the *vestigial* product is central to wing development and seems to be spatially and temporally restricted. The alternative forms of the *vestigial* transcript described herein are often missing the second exon of the previously reported 3.8 kb *vestigial* message. This region possesses the translational start site (WILLIAMS *et al.* 1991) which suggests that alternate transcripts would not encode the Vg protein. Alternative splicing may act as a mechanism by which production of the *vestigial* product is regulated.

The expression of the *vestigial* product is central to wing development and is spatially and temporally restricted. Removing the translational start site by alternative splicing may be a very effective way to control the expression of an important protein. The following model suggests that alternative splicing of the *vestigial* transcript could control *vestigial* expression in two different ways. The developmental pattern of expression of mRNA detected by the *vestigial* probe in Northern analysis reveals no great change in total expression from embryogenesis through pupariation except, perhaps, during the third instar larval stage. In general, the differences appear to be qualitative not quantitative, and result in alternative lengths of mRNA rather than dissimilar amounts of transcripts. During embryogenesis, the first larval instar and pupariation, the longer *vestigial* messages appear to be favoured. The second and third larval stages are dominated by the shorter messages. The adult mRNA detected appears to be different in size and amount from the rest. One level of control could to be the binary choice between the synthesis of long and short transcripts. A second control mechanism could distinguishes between messages within a given size range. The decision to splice out the second exon may determine whether or not the full length 3.8 kb mRNA, and presumably the Vg protein, is produced. The choice between the synthesis of longer messages may be based upon tissue-specific differences. A similar decision may result in the series of smaller transcripts. The differences in message processing may be directly responsible for the pattern of *vestigial* expression for both the complete message and the Vg protein. Clearly, if these are alternate forms of the *vestigial* message, fine tuning of expression may be the result of the expression of proteins involved in the splicing of the pre-mRNA. In future experiments, *vestigial* mutations that influence processing of the message could be characterized. In addition, mutants at loci that contribute to the control of alternative splicing at *vestigial* may be isolated through the analysis of genetic interactions.

The 5' untranslated region of the *vestigial* transcript has three adenine and cytosine (A/C) rich regions, at nucleotide positions 862 1289 and 1454 (WILLIAMS *et al.* 1991). Alternative splicing of *doublesex* (*dsx*) pre-mRNA is regulated by the binding of the Tranformer and Transformer-2 proteins to a splicing regulator downstream of the female-specific splice site (TIAN and MANIATIS 1994). The splicing enhancer is composed of six copies of a A-C rich 13 nucleotide *dsx* repeat element. The element has the consensus sequence of TC(t/a)(t/a)CAATCAACA. The three A/C rich regions in the *vestigial* leader have similar repeats which fit a 16 nucleotide consensus sequence of AA(a/c)AACAACAA(a/c)A(a/c)A(c/a). There are other A-rich stretches present in the second intron that are somewhat similar but do not fit this consensus. These repeats may be candidates for a function similar to the *dsx* repeat element and may be the site of interaction for undetermined RNA-binding factors which may direct the alternative splicing of the *vestigial* message.

Studies of *vestigial* dosage analysis suggest that *vestigial* may control its own expression (GREEN 1946; WILLIAMS 1989). The weak homology of the second conceptual *vestigial* protein to the PTB-associated protein (PATTON *et al.* 1993) suggests that there is a possibility that *vestigial* gene itself may be involved in splicing. One intriguing possibility is, like *Sex-lethal*, the *vestigial* protein itself is involved in the splicing process (BELL *et al.* 1991). This question must await the confirmation of the existence of a hypothetical second *vestigial* protein.

Drosophila RNA binding proteins have been demonstrated to be actively involved in the determination of alternative splicing (BELL *et al*. 1988; MATTOX and BAKER 1991). It is possible that *vestigial* expression is under the influence of unidentified RNA binding proteins. If this model is correct, sequences within the *vestigial* gene are active in the binding of unknown factors which leads to exon-skipping under the appropriate conditions. The A/C rich regions of the second exon and a sequence containing three complete and one partial hexameric repeats of AGTGCC (nt-1725 to nt-1746), also in the second exon, may be involved in RNA-splicing. Considering the essential role that *vestigial* plays in wing development, any gene product that acts to direct its expression may be very important . There may be a number of undiscovered RNA-binding proteins that are essential to the development of the Drosophila wing.

LITERATURE CITED

ARKHIPOVA, I. R., 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. Genetics **139**: 1359-1369.

BELL, L. R., E. M. MAINE, P. SCHEDL and W. T. CLINE, 1988 *Sex-lethal* , a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell **55**: 1037-1046.

BELL, L. R., J. I. HORABIN, P. SCHEDL and W. T. CLINE, 1991 Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in Drosophila. Cell **65**: 229-239.

BOWNES, M., and S. ROBERTS, 1981 Regulative properties of wing discs from the *vestigial* mutant of *Drosophila melanogaster*. Differentiation **18**: 89-96.

CAMPBELL, S., M. INAMDAR, V. RODRIGUES, V. RAGHAVAN, M. PALAZZOLO and A. CHOVNICK, 1992 The *scalloped* gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev. **6**: 367-379.

CAVENER, D. R., 1987 Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acid Res. **15**: 1353-1361.

COHEN, B., M. E. MCGRIFFIN, C. PFEIFLE, D. SEGAL and S. M. COHEN, 1992 *apterous*, a gene required for imaginal disc development in Drosophila encodes a member of the LIM family of developmental regulatory proteins. Genes Dev. **6**: 715-729.

COHEN, S. M., 1993. *Imaginal disc development*, pp. 747-841 in *Drosophila Development*, edited by A. MARTINEZ-ARIAS and M. BATE. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

COHEN, S. M., and G. JURGENS, 1989 Proximal-distal pattern formation in Drosophila: cell autonomous requirement for *Distal-less* gene activity in limb devlopment. EMBO J. **8**: 2045-2055.

FRISTROM, D., 1968 Cellular degeneration in wing development of the mutant *vestigial* of *Drosophila melanogaster*. J. Cell Biol. **39**: 488-491.

FRISTROM, D., 1969 Cellular degeneration in the production of some mutant phenotypes in *Drosophila melanogaster*. Mol. Gen. Genet **103**: 363-379.

GREEN, M. M., 1946 A study in gene action using different dosages and alleles of *vestigial* in *Drosophila melanogaster*. Genetics **31**: 1-20.

HODGES, D., and S. I. BERNSTEIN, 1994 Genetic and biochemical analysis of alternative RNA splicing. Adv. Genet. **31**: 207-281.

MATTOX, W., and B. S. BAKER, 1991 Autoregulation of the splicing of transcripts from the *transformer-2* gene of Drosophila. Genes Dev. **5**: 786-796.

MCKEOWN, M., 1992 Alternative mRNA splicing. Ann. Rev. Cell Biol. **8**: 133-155.

MORGAN, T. H., 1911 The origin of nine wing mutations in Drosophila. Science **33**: 496-499.

NASH, D., and J. B. BELL, 1968 Larval age and the pattern of DNA synthesis in polytene chromosomes. Can. J. Genet. Cytol **10**: 82-92.

O'KEEFE. S., M. SCHOULS and R. HODGETTS, 1995 Epidermal cell-specific quantification of *dopa decarboxylase* mRNA in Drosophila by competitive RT-PCR: an effect of Broad-complex mutants. Develop. Genet. **16**: 77-84.

PATTON, J. G., E. B. PORRO, J. GALCERAN, P. TEMPST and B. NADAL-GINARD, 1993 Cloning and characterization of *PSF*, a novel pre-mRNA splicing factor. Genes Dev. **7**: 393-406.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular cloning: A laboratory manual.* Cold Spring Harbor Press, Cold Spring Harbor, New York.

TALBOT, W. S., E. A. SWYRYD and D. S. HOGNESS, 1993 Drosophila tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. Cell **73**: 1323-1337.

TIAN, M., and T. MANIATIS, 1994 A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev. **8**: 1703-1712.

WILLIAMS, J. A., 1989 The *vestigial* locus of *Drosophila melanogaster.* Ph.D. thesis, University of Alberta.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**:1355-1363.

WILLIAMS, J. A., and S. B. CAROLL, 1993 The origin, patterning and evolution of insect appendages. Bioessays **15**: 567-578.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988a Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988b Suppressible *P*-element alleles of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **212**: 370-374.

WILLIAMS, J. A., A. L. AKTIN and J. B. BELL, 1990 The functional organization of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **221**: 8-16.

WILLIAMS, J. A., J. B. BELL and S. B. CARROLL, 1991 Control of Drosophila wing and haltere development by the nuclear *vestigial* gene product. Genes Dev. **5**: 2481-2495.

WILLIAMS, J. A., S. W. PADDOCK and S. B. CARROLL, 1993 Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. Development **117**: 571-584.

WILLIAMS, J. A., S. W. PADDOCK, K. VORWERK and S .B. CARROLL, 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature **368**: 299-305.

Chapter 6

Insights into the *P* element and the *vestigial* gene

The generation and characterization of *P* element derivatives of *vestigial*^{21} (WILLIAMS and BELL 1988; WILLIAMS *et al.* 1988a;b) is the primary focus of the studies outlined here. These investigations explore the relationships between the *vestigial* gene and the adjacent *P* elements. As an initial step in these studies, a reporter gene subject to the influence of *vestigial* control sequences was desired to study the mechanisms which control *vestigial* transcription. An *in vitro* modified *P* element carrying the bacterial *lacZ* gene (O'KANE and GEHRING 1987; BELLEN *et al.* 1989) inserted into *vestigial* was believed to fulfill this role. However, the low frequency of recovery of *P* element alleles indicated that the isolation of an enhancer trap at *vestigial* would be difficult and inefficient. In consideration of this limitation, a novel technique, namely the targeted transposition of an enhancer trap to this gene, was conceived, designed and executed. Unexpectedly, in contrast to a great number of enhancer trap insertions which faithfully report *lacZ* expression in a manner similar to the detected gene (BELLEN *et al.* 1989), the expression pattern of *vestigial*^{lacZ1} differs from that of the *vestigial* message and protein.

Examination of the structures of internally-deleted *P* elements recovered at the insertion site of *vestigial*^{21} after dysgenic crosses reveal that there is a very strong tendency for *P* elements to become internally deleted with at least one breakpoint located in the inverted repeat. A quite striking clustering of the breakpoints at or near the sixteenth base pair from the end of the transposon suggests that there is a mechanism in place to protect and retain the ends of the *P* element's terminal inverted repeat.

Correlation of the phenotypes of the *vestigial*^{21} derivatives with the nature of the *P* element sequences present at the insertion site indicates that the *P* element expression influences *vestigial* transcription. Analysis of the transcription products of the wild type *vestigial* gene demonstrates that more than the 3.8 kb transcript (WILLIAMS *et al.* 1990a) arise from the *vestigial* locus. The existence of alternative splicing at *vestigial* may provide insight into the mechanisms by which this gene is controlled. The current investigations of *P* element alleles of the *vestigial* gene extend the knowledge of *P* element biology, the *vestigial* locus and their functional interactions.

Targeted enhancer trap transposition: Two similar techniques by which sequences from ectopic locations are copied to the site of a particular *P* element are targeted transposition (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994) and targeted gene replacement (GLOOR *et al.* 1991). In both, *P* element excision is the initial event that produces a double-strand break which is then repaired by DNA synthesis. Targeted transposition replaces one *P* element with another while targeted gene replacement brings sequences homologous to the target site from another site in the genome. The Southern hybridization and DNA sequence analysis of *vestigial*^{lacZ1} demonstrates the enhancer trap has been recruited to the site of the pre-existing *P* element of *vestigial*^{21} . The recruitment of *P[lacZ]* sequences to the *vestigial* gene represents an unique technique as this is the first and only, to date, report of an enhancer trap targeted to a specific gene. This is an extension of previous observations in which *P* elements have been brought to *yellow* (GEYER *et al.* 1988) and in which natural *P* elements (WILLIAMS *et al.* 1988a) and *P* elements containing *dopa decarboxylase* constructs (HESLIP *et al.* 1992; HESLIP and HODGETTS 1994) have been targeted to *vestigial.* The success of this technique implies that it should be possible to target reporter gene constructs to the sites of existing *P* elements located in other genes. This experiment is the prototype of combined targeting and enhancer trap techniques which may be applicable to a great variety of studies.

The potential advantages of enhancer trap targeting are many. Firstly, as in this instance, an enhancer-trap allele of a relative "cold spot" for *P* element insertion is much more likely to be derived from a strain which bears a previously existing *P* element insert. Given the very low frequency of *P* element insertion at *vestigial* (WILLIAMS and BELL 1988), it is unlikely that this locus could have been tagged in a normal screen for enhancer traps. Secondly, reporter gene constructs may be inserted at specific sites within or near the target gene which may react in different manners to specific position effects. Thirdly, specific enhancer trap lines should be isolated at a much higher frequency by this technique than by standard *P* element mobilization methods. Although the cryptic nature of *vestigial*^{21} allows for easy selection of alterations of the resident *P* element, the phenomenon of enhancer trap recruitment may be applied to other genes if a pre-existing *P* element allele is available. This manner of generating enhancer trap alleles may be most applicable to genes in which screening for newly induced mutant alleles is difficult. Alternatively, this procedure may be employed as a mechanism to recruit reporter gene sequences to a specific *P* element insertion site. In addition, this method may make the generation of an enhancer trap allele of a given gene less labour intensive.

Nature of the double strand gap: *P* element transposition acts through a cut-and-paste mechanism and must occur in a manner that generates an eight base pair target site duplication (KAUFMAN and RIO 1992). The current model of transposition (SHAPIRO 1979) suggests that for such a duplication to arise, a staggered cut must be made at the target to generate 5' overhanging termini to which the transposon's end can ligate.

Synthesis initiated at the recessed 3' end fills in the complementary strand to produce the duplication. Since 5' overhanging ends on the transposon would not result in the target site duplication described, the nature of the excisions flanking the transposon is restricted to leaving either blunt or 3' overhanging ends. The nature of the cut has been investigated in the embryo by examining the effects of transposition from a plasmid bearing a *P* element (O'BROCHTA *et al.* 1991). Pre-blastoderm embryos were co-injected with a plasmid containing a *P* element cloned into *lacZ* and a plasmid bearing the transposase gene. The former was later recovered and those carrying excisions were examined. Two of the most common derivatives isolated retained either one of the two similar sequences CATGATG or CATCATG at the plasmid's donor site. These appear to be the terminal seven bases of one end or the other of the *P* element. The palindromic terminal seven base pairs are inverted repeats and are also a 6/7 match for a direct repeat. The other common plasmid recovered has the sequence CATG at the donor site. This is also a palindrome. The sequences present at the excision site are consistent with staggered cuts which leave overhangs of 4 to 7 base pairs that anneal and, in the first two classes, resolve the central C-C pair to either G-C or C-G. A second possibility is that the staggered cuts leave longer overhangs that anneal at either the last four or seven base pairs and resolve by trimming away the remaining single stranded tails. It should be apparent that *P* elements excised from plasmids injected into embryos may not be limited by the constraints of *P* elements in the germline.

When the *P* element of *vestigial*^{21} is excised during the formation of germ cells, a different spectrum of products is recovered. Although a number of asymmetric internally deleted *P* elements were recovered, no plasmids were isolated that had apparently undergone invasion of another template in the experiments discussed above. The internally deleted *P* elements most often have 16±1 base pairs of the terminal repeats retained at one terminus or both. This is consistent with transposition leaving behind an overhang of approximately 16 nucleotides. The smallest *P* elements, roughly a mirror image duplication of the terminal 16±1 base pairs, appear to result from two free ends abutting and then fusing. The differences between the *P* element structures obtained at *vestigial*^{21} and those from the plasmid assay are quite apparent. The differences may be either due to the specification of the dimensions of the staggered cut during excision or to protection of the *P* element ends by binding of a protein to the ends or both. This may reflect another fundamental difference between the somatic and germline assays. The inverted repeat binding protein may thus determine the nature of the double strand gap.

Terminal internal deletion of the *P* element: The tendency for *P* elements to become internally-deleted such that at least one deletion breakpoint is situated within the terminal inverted-repeats is common (SEARLES *et al.* 1986; WILLIAMS *et al.* 1988a; TAKASU-ISHIKAWA *et al.* 1992; JOHNSON-SCHLITZ and ENGELS 1993). A synthesis-dependent strand annealing (SDSA) model was proposed to explain gene targeting at the *white* gene (NASSIF *et al.* 1994). According to this model, the free ends left after excision invade a homologous sequence located on the sister chromatid, the homologue or at an ectopic site. Both ends initiate DNA synthesis and copy the template. The process is completed by the annealing of the two strands in a complementary region and continued DNA synthesis after this template switch until the single stranded regions are filled in. If an ectopic template is utilized, sequences are recruited to the site. If the sister chromatid is copied, the original *P* element may be restored. This mechanism is consistent with the unidirectional flow of information from the donor sequence to the target which causes changes only in the recipient. If the DNA synthesis is interrupted, annealing may take place at small regions of homology to give rise to an internally deleted *P* element. The sites of internal deletion within *P* elements are often flanked by sequences which have short direct repeats at both sides (ENGELS *et al.* 1990). Modified *P* elements harbouring long stretches of direct repeats are preferentially deleted such that one repeat and the intervening material is removed when a transposase source is provided (KURKULOS *et al.* 1994). These findings suggest that the SDSA model may also account for many internally deleted *P* elements.

The nature of internally-deleted *P* elements at *vestigial* and elsewhere suggest that the symmetric invasion of each side of a double-strand break is not the only way to repair the gap left by transposition. As described above, the gap can be repaired by the two free ends on either side of the double strand gap abutting and joining. In the case of the very smallest *P* elements recovered, those that retain only 16 or so base pairs of each terminus fused together, the simplest model to explain their origin involves the two 3' overhanging ends abutting or overlapping by a base pair or two, followed by DNA synthesis to cross the gap. The IRBP has the potential to mediate this activity by maintaining the single-stranded 3' overhang, thus preserving a template to be copied. If only one of the two ends invades, the copied material can be included at the insertion site by the strand switching templates from the one that is being copied to the other free end protected by the IRBP. Either end can initiate synthesis since we have isolated terminal internal deletions which break in either the 5' or the 3' repeat. Not only can this explain the internally deleted *P* elements at *vestigial* but may also account for the structures of transposons recovered at *white* which have a truncated *yellow* gene fused to the last 14 or 17 base pairs of the *P* element inverted repeat (NASSIF *et al.* 1994). This is a DNA synthesis mechanism that does not depend upon donor transposition, so if DNA synthesis carries through the donor *P* element sequences adjacent to the *P* element can be brought to the target site. The recovery of composite *P* elements at *vestigial* (HESLIP *et al.* 1992) and *rudimentary* (TSUBOTA and DANG-VU, 1991) suggest that flanking DNA sequences were recruited from a donor site by DNA synthesis that continued through the ectopic *P* element and into neighbouring sequences to be finally resolved by switching to the protected template. The structure of the ends of the *vestigial*^{lacZ1} transposon reveals a precise replacement at one end and a complex sequence at the other end which seems to reflect a template switch within the terminal repeat of the donor *P* element, a stretch of filler sequence and the terminal 16 base pairs of the *P* element.

Among the internally-deleted derivatives of *vestigial*^{21} that are missing 3' *P* element sequences, template switching seems to favour regions located immediately upstream of AT rich sequences. This is consistent with the possibility that the processivity of this DNA synthesis is sensitive to the base composition of the donor element, in this case the *vestigial*^{21} *P* element. In essence, either the copying of the template appears to be interrupted at a point that precedes a stretch of AT rich sequences or, once copied. they may tend to preferentially interact with the *P* element terminus during resolution of the gap. In the first circumstance, the newly synthesized AT rich single strand may be prone to dissociation from the template. The dislodged sequence may then be shortened by proofreading before synthesis is continued using the protected terminus as a template. The terminal sequence has an AT rich nature which may have a stronger tendency to interact with the AT rich regions than with other portions of the newly synthesized single strand. Possibly, both activities are involved in the resolution of an asymmetric invasion event, such that disassociation from the template initiates the switching from the donor to the protected terminal sequence.

Replication slippage during DNA repair: The filler sequences sometimes located at the deletion breakpoints of internally-deleted *P* elements have been suggested to originate via a replication mechanism that utilizes a hairpin intermediate (TAKASU-ISHIKAWA *et al.* 1992). In this model, the free ends would be repaired by DNA synthesis turning in on the same strand to generate a loop structure followed by resolution through breakage and rejoining of the two hairpin ends. The filler sequences found at *vestigial* are often short simple tandem repeats identical to sequences present in the *P* element termini. This is consistent with a model of replication and slippage, originally suggested by O'HARE and RUBIN (1983), using sequences from the inverted-repeat as a template. The non-repeat AT rich filler sequences found in these studies may also be generated in this way. The frequency of *P* elements which possess filler sequences appears to be high among those that have at least one breakpoint in the inverted repeats. The switching of templates from the donor sequence in an attempt to bridge the gap appears to often involve replication slippage. Perhaps the repair of double-strand breaks caused by either the transposition of mobile elements or by DNA damage may often lead to the creation of filler sequences to produce small sequence polymorphisms. This process may be similar to that which leads to unstable microsatellites implicated in several human genetic diseases (KUNKLE 1993).

Role of the inverted-repeat binding protein (IRBP): According to current models, *P* element transposition leaves behind a double-strand break (ENGELS 1989; ENGELS *et al.* 1990; KAUFMAN and RIO 1992). The transposase binds to sites internal to the terminal repeats (KAUFMAN *et al.* 1989; KAUFMAN and RIO 1991) but produces cuts such that at least one strand is cleaved at each of the *P* element's edges. This suggests that other protein factor(s) bind to the terminal repeats to specify the site of the cleavage reaction. The inverted repeat binding protein (RIO and RUBIN 1988; BEALL *et al.* 1994), due to its ability to bind the outer half of the 31 base pair inverted repeat sequence, appears to be a very good candidate for this activity. In addition, in order for the double-strand break caused by transposition not to result in chromosome loss, the gap must be efficiently repaired. If, theoretically, the processes of transposition and gap repair were linked, perhaps sharing a common protein component, the survivorship of dysgenic gametes should be higher than if the activities were not.

The locus encoding the IRBP has been cloned recently by screening a cDNA expression library with anti-IRBP antibodies followed by isolation of full-length embryonic cDNA clones (BEALL *et al.* 1994). The sequence of IRBP is homologous to the human 70-kDa Ku protein subunit (REEVES and STHOEGER 1989) that forms a heterodimer with a 86-kDa protein (YANEVA *et al.* 1989) to constitute the human Ku antigen. The Ku antigen was originally identified as a DNA-binding protein present in the sera of patients of autoimmune diseases such as scleroderma and lupus erythematosus. A single 2.4 kb IRBP mRNA is expressed at low levels throughout development but is present at much higher levels in females, ovaries and early embryos. This suggests that the highest levels of expression occur during embryogenesis.

Cytogenetic mapping of the IRBP gene places it at polytene chromosome region 86E2-3. A member of a class of mutagen sensitive genes (BOYD *et al.* 1981), *mus309*, has been mapped to or near this region (LINDSLEY and ZIMM 1992). The sterility associated with *mus309* alleles is consistent with the expression pattern of IRBP mRNA in germline. The IRBP may be encoded by *mus309* (D.C. RIO, personal communication) but this has yet to be confirmed .

The inverted repeat binding protein is not encoded by the *P* element, but is a host-derived protein that has likely been usurped into a role involved in the transposition of the *P* element (RIO and RUBIN 1988; BEALL *et al.* 1994). The apparent direction of the transposase by the IRBP to make a staggered cut to produce a 5' overhang of 16 or so bases seems to be its novel role. DNA repair may be the normal primary duty of IRBP. This is consistent with mutagen sensitivity and some proposed activities of the Ku antigen. An attractive idea is that the transposase replaces the 86 kDa protein and forms a complex which utilizes the DNA binding ability to result in transposition instead of repair. The remaining free ends, occupied by proteins important in repair, may be readily joined.

The activity of IRBP in the repair of double-strand DNA gaps generated by *P* element excision proposed above may provide a clue to its role in the absence of *P* elements. Some of the repeat structures generated as filler sequences during post-excision repair are similar to heterochromatic satellite sequences. The junctions between tandem arrays of satellite sequences in *Drosophila melanogaster* seem to switch from one tandem repeat to another such that the periodicity is maintained (LOHE and BRUTLAG 1987). To account for the origin of a smooth transition between two similar satellite sequences, a model of template amplification was presented. According to this mechanism, during the synthesis of an array of satellite DNA each repeat is copied using the previous repeat as a template such that errors in replication that result in the inclusion of a base change into the array leads to a smooth transition. The amplification of satellite sequences may be a response to DNA damage and thus may be similar to the repair of *P* element induced gaps. The IRBP may be widely involved in DNA synthesis-dependent repair strategies to fix double strand gaps.

***Cis* regulation of *vestigial* by the *P* element:** The set of transposase-induced deletion derivatives of the *P* element of *vestigial*^{21} provides indications of the molecular basis for the effect of the *P* elements upon the expression of *vestigial* . The founder allele, *vestigial*^{21}, is a weak allele with a cryptic phenotype caused by the insertion of an incomplete *P* element. All of the phenotypic revertants have deletions of the 5' region of the *P* element. In contrast, all the strong *vestigial* alleles derived from *vestigial*^{21} retain at least the first 105 base pairs. The promoter and the transcriptional start site of the *P* element transcription unit are present in this 5' region (KARESS and RUBIN 1984). Also, the polyadenylation site of the *P* element transcript (LASKI *et al*. 1986) is absent in all of these alleles. Sequence analysis of the extreme derivatives demonstrates that the presence of the *P* element promoter without the polyadenylation signal results in a *vestigial* phenotype. The absence of the *P* element transcriptional start site apparently returns *vestigial* expression to wild type levels even if a large amount of the *P* element exists at the insertion site. The *vestigial* phenotypes of these mutants are greatly influenced by the presence or absence of *P* element control sequences.

Two features of the *P* element are essential in the interpretation of the mutant phenotypes and the molecular lesions associated with the *P* element alleles of *vestigial*. First, the *P* element promoter is active in somatic tissues and second the transcription of the *P* element can read through its polyadenylation signals and into adjacent sequences (LASKI *et al*. 1986). The correlation between the primary sequence and the resultant phenotype is consistent with the possibility that transcription initiated at the *P* element promoter reduces *vestigial* gene function. The transcriptional interference may be central to the control of the *vestigial* locus by the *P* element sequences. Initiation of transcription at the *P* promoter could act to compete RNA polymerase and transcription factors away from the downstream *vestigial* promoter. The *P* element promoter (KARESS and RUBIN 1984) is a better match to the consensus TATA box than the *vestigial* promoter. The process of transcribing through the *vestigial* promoter may be sufficient to interfere with the initiation of legitimate *vestigial* transcription. The precise mechanism by which the interference is accomplished remains unknown.

The existence of *P* element-gene fusions that result in decreased expression is consistent with the behaviour of the cytotype dependent alleles of *vestigial* (WILLIAMS *et al*. 1988b) and *singed* (ROBERTSON and ENGELS 1989). The *vestigial*^{21} allele is somewhat sensitive to *P* element cytotype and displays complete wings as a heterozygote with a null allele instead of strapped wings (WILLIAMS *et al*. 1988b). The most sensitive allele, *vestigial* $^{21\text{-}3}$, has the classic *vestigial* reduction of wing tissue in the M cytotype but develops normal wings in the *P* cytotype, and has been utilized as an assay to detect *P* element repressor activity (RIO 1990; GLOOR et al,. 1993). If *P* element repressor molecules bind at the 5' binding site that overlaps with the *P* element promoter (KAUFMAN *et al.* 1989; KAUFMAN and RIO 1991), transcription of the *P* element-*vestigial* fusion transcript may not be initiated. The binding of the repressor to other site(s) in the *P* element may also contribute to reducing transcriptional read through. Thus, little interference of the normal *vestigial* transcript need occur. As a result, *vestigial* expression may return to levels approaching wild type in a *P* cytotypic background. This may be similar to the effect of generating *P* element revertant alleles that do not retain the *P* element promoter. The control of *vestigial* expression by the *P* element protein, appears to be somewhat analogous to the selection of alternative promoters dependent upon the presence or absence of a transacting factors which selects one promoter or another under different conditions.

Fusions of the *P* element and *vestigial* sequences result in recessive alleles of *vestigial* because the major consequence of the hybrid message seems to be reduction of the normal message. Gene fusions of a different nature have been previously identified at *vestigial*. The two dominant *vestigial* alleles, $vestigial^{W}$ and $vestigial^{U}$, are both the result of chromosomal inversions that cause the juxtaposition of *vestigial* and two other genes, *mastermind* and *invected* (WILLIAMS *et al.* 1990b). WILLIAMS concluded that since the transcript units were no longer intact the dominance was probably not caused by inappropriate expression of *vestigial*. Recently, reversion of the dominant lack-of-wing phenotype of $vestigial^{W}$ demonstrated that its dominant homeotic phenotype, an anterior to posterior wing transformation, is separate from the dominant no-wing phenotype (KOVITHATHONGS and BELL, unpublished). In light of this finding and the elucidation of *vestigial* as a selector gene in the specification of the wing margin (Williams *et al.* 1994), ectopic expression of the truncated *vestigial* transcripts driven by the control sequences of *mastermind* ($vestigial^{U}$) or either *invected* and/or *engrailed* ($vestigial^{W}$) is quite probably the cause of the dominant phenotypes. The *P* element-*vestigial* fusion differs from the dominant gene fusions because it seems to direct the inappropriate inhibition of the *vestigial* transcript.

Implications for molecular evolution: The mechanism that may be responsible for the phenotypic effects of $vestigial^{21}$ and its derivatives is the appropriation, by the adjacent *P* element, of the control of *vestigial* transcription. The revertant alleles are relieved of the effects of the gene fusion and the extreme alleles have the expression of *vestigial* greatly decreased by the action of the transposon's promoter. Cytotype dependent alleles, such as $vestigial^{21-3}$, exhibit levels of *vestigial* expression that are either relatively free of the effects the transposon's promoter or are greatly influenced by it depending upon the presence or absence of the trans-acting *P* repressor. The capture of a coding region by a transposable element's control region may be a common source of genetic variation.

A single repetitive array of degenerative *P* elements has been isolated in the species *Drosophila gauche* (MILLER *et al.* 1992). Regions of exons 0 1 and 2 are similar to the *P* element of *Drosophila melanogaster*, although the rest of the *P* element sequence has diverged significantly. The *P* homologues are not mobile because the inverted repeats and the transposase-specific third exon are not present. However, the potential ability to encode a protein that is very similar to the *P* element repressor protein is maintained, which suggests that it may have adopted a regulatory role in this species. The function of the degenerative *P* element and the elucidation of the targets of the *P* protein will be of interest and may demonstrate how the genes of a transposable element can become integrated into the genetic makeup of the host.

Transcriptional start site: The site at which transcription of the *vestigial* message is initiated has been mapped at a position 210 base pairs upstream of the splice donor site of the first exon. The insertion site of the *vestigial*^{21} *P* element is 0.5 kb upstream of this location. The large fusion transcript includes genomic sequences in the intervening region not normally part of the *vestigial* transcript which, in addition to the *P* element sequences, may contribute to altered expression.

Alternative splicing at *vestigial*: In alternative splicing, the primary transcripts are cleaved and rejoined in different ways to produce more than form of mature mRNA (MCKEOWN 1992; HODGES and BERNSTEIN 1994). The tendency of a gene to produce different forms of message can be considered a method to control gene expression. The mechanisms that control splicing and alternative splicing are currently receiving much attention and much has been determined recently about the processes that lead to alternatively spliced transcripts. *Drosophila melanogaster* is rapidly proving to be an excellent organism in which to study these processes because of the great availability of genetic, biochemical and molecular techniques in this model organism.

Strategies of alternative splicing can be sub-divided into several classes of molecular decisions. Splicing may simply occur or not occur. The decision to retain a section of RNA which is often removed may introduce a premature stop codon to produce a truncated protein. This is the mechanism by which the *P* element transcript is spliced in the germline to produce a message which encodes the transposase and is not spliced in the soma to produce the smaller repressor molecule (LASKI *et al.* 1986; SIEBEL and RIO 1990). Alternate splice sites may lengthen or shorten the mRNA to include or remove sequences such as the exclusion of an early stop codon in the smaller female-specific message of *transformer* (BOGGS *et al.* 1987). Exon skipping, the exclusion of one or more of the internal exons, removes sequences present in the longest spliced version of the message. The *Sex-lethal* locus is initially activated by a female embryo-specific alternative promoter to produce Sex-lethal protein which directs the exclusion of an exon present in the nonfunctional male to produce more Sex-lethal protein in the female (BELL *et al.* 1988; 1991). The various approaches in alternative message processing act to control expression of the protein products of the genes.

Previous examination of the expression of *vestigial* concentrated on one transcript which hybridized to several *vestigial* probes. The other transcripts, some of which apparently demonstrated a very strong affinity for at least one of the probes used, were attributed to artefactual cross hybridizations. This is not consistent with the additional products isolated in this work. The existence of alternative splicing at *vestigial* may provide insights into the means by which expression of the protein is controlled. If most of the bands observed in the Northern analysis are actually different forms of the transcript then it appears that two general strategies may be invoked in the control of *vestigial* expression. During the embryonic stages and again during the pupal stage longer transcripts are dominant on the developmental Northern blots. During the interim larval stages shorter transcripts are prevalent. This is consistent with a mode of control which favours smaller transcripts during the larval stages and larger ones during embryogenesis and pupariation. Another strategy appears to differentiate between messages that contain the second exon and the smaller ones, perhaps in a tissue specific manner. The bands produced in the Northern analysis appear to change in a qualitative not quantitative manner way during development. The total amount of message detected with the 3' *vestigial* probes does not seem to vary to the extent that a developmentally regulated gene would suggest. However, the presence of the longest mRNA is consistent with such expression.

The 3.8 kb message can code for the nuclear *vestigial* protein (WILLIAMS *et al.* 1991). The translational start site is located within the second exon which is spliced out in at least some of the alternative transcripts. Exclusion of the second exon quite possibly results in a *vestigial* mRNA which does not produce a functional protein product. The exclusion of the start codon to produce a nonfunctional mRNA is opposite from the strategies used by *Sex-lethal* (BELL *et al.* 1988; 1991) and *transformer* (BOGGS *et al.* 1987) which produce non-functional messages when stop codons are included in the final transcript. Another transcript which splices out the normal translational start site is the ecdysone receptor isoforms B1 and B2 (TALBOT *et al.* 1993). When the second exon of the B1 isoform is spliced out, the B2 mRNA is produced which does not possess the B1 translation start site but utilizes an otherwise unused start codon located upstream. The approach to control translation which *vestigial* appears to use is unusual and may be an example of a as yet unreported post-transcriptional regulatory mechanism.

The identification of alternative splicing at *vestigial* suggests that further analysis may reveal the mechanisms by which this is controlled. The sex determination pathway was characterized through genetic methods before this cascade of regulated RNA binding proteins was uncovered. The *vestigial* gene may function extremely well as a model system to isolate and analyze splicing components. As a well characterized, non-vital gene whose major effect is easily observed in the adult wing structures, it may be particularly sensitive to transacting factors which influence splicing. There exists a large number of *vestigial* mutations and genes that modify *vestigial* expression. Examination of splicing of transcripts from various *vestigial* alleles in different genetic backgrounds and the isolation of second-site suppressor and enhancer mutations may provide valuable clues to the mechanisms controlling alternative splicing at *vestigial*.

Reporter gene activity of *vestigial*^{lacZ1}: In the imaginal wing discs of third instar larvae, the nuclear-localized *vestigial* protein is expressed to its maximum level along the length of the presumptive wing margin, decreasing in sharp gradients on both the dorsal and ventral side of the margin (WILLIAMS *et al.* 1991; 1993). A wing and haltere disc specific enhancer which apparently drives this expression is present within the second intron (WILLIAMS *et al.* 1994). The wing and haltere imaginal discs of vg^{lacZ1}/vg^{+} larvae were examined for ß-galactosidase activity and expression in the wing disc was found to be localized in a region in the centre of the disc which corresponds to presumptive dorsal wing surface near the wing hinge (fate map: BRYANT 1975). Haltere expression appears to be localized a homologous region and imaginal leg discs do not show any activity. In contrast to many enhancer trap alleles which express lacZ in the same manner as the affected gene (BELLEN *et al.* 1989), the wing and haltere expression of ß-galactosidase in *vestigial*^{lacZ1} does not mirror the mRNA and protein expression of *vestigial* in the presumptive wing margin (WILLIAMS *et al.* 1991). Wing and haltere specific controls do act upon the lacZ construct, but as the difference between normal *vestigial* expression and the reporter gene expression indicates, this transgene is not apparently influenced by the mechanisms that limit expression to the presumptive wing margin. It is possible that the wing and haltere expression may represent the persistence of ß-galactosidase activity driven by *vestigial* control elements active in an earlier stage of development. Alternatively, the observed reporter gene activity, which may mirror early transient *vestigial* expression, has not been discontinued due to the failure of control elements, which normally down regulate *vestigial* expression, to act upon this construct. Thus, this reporter gene may be revealing a previously unidentified subset of wing and haltere control element(s).

The expression of *vestigial*^{lacZ1} is not identical to the *vestigial* mRNA or protein. Even so, this transgene may be responding to a positional cue or morphogenic signal which may be very important in wing development. The site within the wing disc where *vestigial*^{lacZ1} expresses ß-galactosidase activity bears further investigation. In the investigation of duplication and regeneration of surgically bisected wing discs, the presence of a central region of the presumptive dorsal wing surface in a cultured wing explant led to regeneration of the disc fragment and its absence to duplication (BRYANT 1975). Eventually these and other experiments led to the proposal of the 'polar-coordinate model' of pattern formation (BRYANT *et al.* 1981). Expression of ß-galactosidase in *vestigial*^{lacZ1} correlates very well with the region of wing that led to regeneration in the above experiment. However, the importance of this correlation, if any, remains to be established. Several reports have implicated wingless (wg, DWnt-1) as an important signal in the establishment of the polar coordinate system in imaginal discs (COUSO *et al.* 1993; STRUHL and BASLER 1993; WILLIAMS *et al.* 1993). This activity may be crucial in the role of *vestigial* as a global wing determinant.

Recently, the investigation of the *rotund* (AGNEL *et al.* 1992) and *nubbin* (NG *et al.* 1995) genes has suggested that a hinge specific organizing centre may exist. The *rotund* gene is expressed in the central region of all major imaginal discs (AGNEL *et al.* 1992). In the wing disc, *rotund* is expressed throughout the wing pouch and adjacent folds and in two small areas, a band in the presumptive dorsal hinge and a proximal patch. The *rotund* mutant wing phenotype appears to be disrupted in the region distal to the hinge (KERRIDGE and THOMAS-CAVALLIN 1988). In *nubbin* mutants, the pattern of *wingless* expression in the third instar wing disc is altered such that the staining along the dorsal fold is significantly reduced (NG *et al.* 1995). The morphology of the wings of *nubbin* mutants also reveals disruptions near the hinge region. The region of the wing disc affected in both *rotund* and *nubbin* mutants corresponds with the central region of staining observed in the *vestigial*^{lacZ1} line. The central disc staining is consistent with a response to the hinge specific organizing region. However, since no distinct border specific morphological structures are associated with this region, the existence of an organizing centre is more likely to be involved in the establishment of early positional values. The earliest known markers for cells that assume wing and haltere disc fates are *vestigial* and *scalloped* (WILLIAMS *et al.* 1993). Re-initiation of this organizing principles may be the molecular basis of the positional values required for wing formation during regeneration. The central signal, to which the reporter gene of the targeted construct may be responding, and the signal which reorganizes the wounded blastonema may be the same.

The role of *vestigial*: The precise function of *vestigial* still remains unclear. At the wing margin, *vestigial* acts to determine the edge. The activation of the *vestigial* second intron enhancer along the dorsal/ventral boundary (WILLIAMS *et al.* 1994) suggests that the *vestigial* gene product is required for growth of the wing pouch and perhaps acts to determine the position of the future wing margin. The *vestigial* and *scalloped* genes may act together as transcriptional regulators of a signaling pathway at the presumptive wing margin and both could assign positional values by regulating patterning pathways. The *scalloped* gene product is a highly conserved TEA-domain DNA-binding protein which may require interaction with a tissue specific protein to activate transcription (CAMPBELL *et al.* 1992). The *vestigial* protein is nuclear and may possibly be the transcription regulator with which *scalloped* interacts. The cells at the margin may then produce a diffusible morphogen or start a series of inductive signals from the boundary to stimulate growth and differentiation of the wing (WILLIAMS *et al.* 1994).

Formation of the dorsal/ventral boundary as a linear growth organizer may be the key difference between the sheet versus tube appendage (COHEN 1993; WILLIAMS and CARROLL 1993). The differences in morphology between the dorsal flight appendages and the legs is likely due to the definition of two different types of focal points in the development of a proximal/distal axis. These are a linear distal margin in the wing and haltere discs and a point distal focus in the leg discs. The *vestigial* gene product is one of the earliest wing-specific determinants (WILLIAMS *et al.* 1991) although the wing margin does not seem to be determined until late third instar. Its early activity may be analogous to the determination of a central focus in the ventral discs which may be the activity that *vestigial*^{lacZ1} is reporting. The point focus may be equivalent to the hinge specific organizer that the activities of *rotund* and *nubbin* seem to reflect. This is consistent with an early role for *vestigial* in the specification of the wing disc. Also, the primary focus may be responsible for the overall organization of the wing disc, which may be reactivated in the regeneration of cultured wing disc fragments (BRYANT 1975). Reiteration of an early point organizer as a linear focus at the presumptive wing margin may be the mechanism by which a sheet-like appendage is developed. Activation of an early focus to establish wing identity followed by re-activation of an signal to establish wing margin is consistent with two organizing centres at the hinge and along the wing margin.

Future prospects: The work presented here, for the most part, has contributed to both the understanding of the *vestigial* gene and the *P* element transposon. As is often the case, many questions have been left unanswered. What is the relationship between the reporter gene expression pattern observed in the *vestigial*^{lacZ1} allele and normal *vestigial* expression? Is *mus309* the locus encoding the IRBP and can *mus309* mutant alleles influence the processes that result in internal deletion of *P* elements? What other genes may be involved in the repair of double-strand gaps left by transposition? How do the interactions between *vestigial* and the *P* element influence the life of the mutants? What sequences contribute to the alternative splicing of the *vestigial* transcript? Do any of the existing *vestigial* mutants result from defects in alternative splicing? What aspects of biology revealed herein have general implications? The continuation of the investigiations described here show great promise in addressing a wide range of biological questions.

In summary, the experiments performed during the course of these studies and the conclusions that were drawn extend our understanding of the *vestigial* gene, *P* element biology and the interaction of one upon the other. First, a novel technique, the targeted transposition to *vestigial* of a reporter gene-bearing *P* element is described. The pattern of *lac*Z gene activity in the targeted line differs from real *vestigial* expression and may respond to signals consistent with a proposed organizing focus located in the centre of the wing disc. During the screen to recover *vestigial*^{lacZ1} a number of deletion derivatives of *vestigial*^{21} were isolated. The great tendency of *P* elements to become internally deleted with one or both breakpoints located 16 base pairs from their termini became apparent through the characterization of derivatives of *vestigial*^{21}. The inverted-repeat binding protein (IRBP) may have a substantial role in the process that leads to internal deletion through a mechanism which is a variant of the synthesis-dependent strand annealing model. The presence or absence of particular *P* element sequences in the different classes of *vestigial* ^{21} derivatives demonstrates that *P* element transcriptional activity influences the expression of *vestigial* . Re-examination of *vestigial* transcription products reveals that the *vestigial* pre-mRNA seems to undergo alternative splicing. As a result, the expression of *vestigial* may be very sensitive to the mechanisms which control differential mRNA processing. As a whole, the set of investigations presented here contribute to the understanding of *P* element biology, examine the interactions of a transposon and a host gene and further the study of an important developmental gene which was initiated in the laboratory of Thomas Hunt Morgan.

LITERATURE CITED

AGNEL, M., L. RODER, R. GRIFFIN-SHEA and C. VODA, 1992 The spatial expression of Drosophila *rotund* gene reveals that the imaginal discs are organized in domains along the proximal-distal axis. Roux's Arch. Dev. Biol. **210**: 284-295.

BEALL, E .L., A. ADMON and D. C. RIO, 1994 A Drosophila protein homologous to the human p70 Ku autoimmune antigen interacts with the *P* transposable element inverted repeats. Proc. Natl. Acad. Sci. USA **91**: 12681-12685.

BELL, L. R., E. M. MAINE, P. SCHEDL and W. T. CLINE, 1988 *Sex-lethal*, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell **55**: 1037-1046.

BELL, L. R., J. I. HORABIN, P. SCHEDL and W. T. CLINE, 1991 Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in Drosophila. Cell **65**: 229-239.

BELLEN, H. J., C. J. O'KANE, C. WILSON, U. GROSSNIKLAUS, R. K. PEARSON and W. J. GEHRING, 1989 *P*-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev. **3**: 1288-1300.

BOGGS, R. T., P. GREGOR, S. IDRISS, J. M. BELOTE and M. MCKEOWN, 1987 Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell **50**: 739-747.

BOYD, J. B., M. D. GOLINO, K. E. S. SHAW, C. J. OSGOOD and M. M. GREEN, 1981. Third-chromosome mutagen-sensitive mutants of *Drosophila melanogaster*. Genetics **97**: 607-623.

BRYANT, P. J., 1975 Pattern formation in the imaginal wing disc of *Drosophila melanogaster*: Fate map, regeneration and duplication. J. Exp. Zool. **193**: 47-78.

BRYANT, S. V., V. FRENCH and P. J. BRYANT, 1981 Distal regeneration and symmetry. Science **212**: 993-1002.

CAMPBELL, S., M. INAMDAR, V. RODRIGUES, V. RAGHAVAN, M. PALAZZOLO and A. CHOVNICK, 1992 The *scalloped* gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev. **6**: 367-379.

COHEN, S. M., 1993. *Imaginal disc development*, pp. 747-841 in *Drosophila Development*, edited by A. MARTINEZ-ARIAS and M. BATE. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

COUSO, J. P., M. BATE, and A. MARTINEZ-ARIAS, 1993 A *wingless*-dependent polar coordinate system in Drosophila imaginal discs. Science **259**: 484-489.

ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 437-484 in *Mobile DNA.*, edited by D.E. BERG and M.M. HOWE. American Society for Microbiology, Washington DC.

ENGELS, W. R., D. M. JOHNSON-SCHLITZ, W. B. EGGLESTON and J. SVED, 1990 High-frequency *P* element loss in Drosophila is homolog dependent. Cell **62**: 515-525.

GEYER, P. K., K. L. RICHARDSON, V. G. CORCE and M. M. GREEN, 1988 Genetic instability in *Drosophila melanogaster*: *P*-element mutagenesis by gene conversion. Proc. Natl. Acad. Sci. USA **85**: 6455-6459.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. R. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in Drosophila via *P* element-induced gap repair. Science **253**: 1110-1117.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS, W. K. BENZ, H. M. ROBERTSON and W. R. ENGELS, 1993 Type I repressors of *P* element mobility. Genetics **135**: 81-95.

HESLIP, T. R., and R. B. HODGETTS, 1994 Targeted transposition at the *vestigial* locus of *Drosophila melanogaster*. Genetics **138**: 1127-1135.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131**: 917-927.

HODGES, D., and S. I. BERNSTEIN, 1994 Genetic and biochemical analysis of alternative RNA splicing. Adv. Genet. **31**: 207-281.

JOHNSON-SCHLITZ, D. M., and W. R. ENGELS, 1993 *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila melanogaster*. Mol. Cell Biol. **13**: 7006-7018.

KARESS, P. D., and G. M. RUBIN, 1984 Analysis of *P* transposable element functions in Drosophila.. Cell **38**: 135-146.

KAUFMAN, R. E., and D. C. RIO, 1991 Drosophila *P*-element transposase is a transcriptional repressor *in vitro*. Proc. Natl. Acad. Sci. USA **88**: 2613-2617.

KAUFMAN, R. E., and D. C. RIO, 1992 *P* element transposition *in vitro* proceeds by a cut and paste mechanism and uses GTP as a cofactor. Cell **69**: 27-39.

KAUFMAN, R. E., R. F. DOLL and D. C. RIO, 1989 Drosophila *P* element transposase recognizes internal *P* element DNA sequences. Cell **59**: 359-371.

KERRIDGE, S., and M. THOMAS-CAVALLIN, 1988 Appendage morphogensis in Drosophila: a developmental study of the *rotund* (*rn*) gene. Roux's Arch. Dev. Biol. **197**: 19-26.

KUNKLE, T. A., 1993 Slippery DNA and diseases. Nature **365**: 207-208.

KURKULOS, M., J. M. WEINBERG, D. ROY and S. M. MOUNT, 1994 *P* element-mediated *in vivo* deletion analysis of *white-apricot*: deletions between direct repeats are strongly favored. Genetics **136**: 1001-1011.

LASKI, F. A., D. C. RIO and G. M. Rubin, 1986 Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44**: 7-19.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

LOHE, A. R., and D. L. BRUTLAG, 1987 Adjacent satellite DNA segments in Drosophila. Structure of junctions. J. Mol. Biol. **194**:171-179.

MCKEOWN, M., 1992 Alternative mRNA splicing. Ann. Rev. Cell Biol. **8**: 133-155.

MILLER, W. J., S. HAGEMANN, E. REITER and W. PINSKER, 1992 *P*-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. Proc. Natl. Acad. Sci. **89**: 4018-4022.

MORGAN, T. H., 1911 The origin of nine wing mutations in Drosophila. Science **33**: 496-499.

NASSIF, N., J. PENNEY, S. PAL, W. R. ENGELS and G. B. GLOOR, 1994 Efficient copying of nonhomologous sequences from ectopic sites via *P* element-induced gap repair. Mol. Cell. Biol. **14**: 1613-1625.

NG, M., F. J. DIAZ-BENJUMEA and S. M. COHEN, 1995 *nubbin* encodes a POU-domain required for proximal-distal patterning in the Drosophila wing. Development **121**: 589-599.

O'BROCHTA, D. A., S. P. GOMEZ and A. M. HANDLER, 1991 *P* element excision in *Drosophila melanogaster* and related drosophilids. Mol. Gen. Genet. **225**: 387-394.

O'HARE, K., and G. M. RUBIN, 1983 Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**: 25-35.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84**: 9123-9127.

REEVES, W. H., and Z. M. STHOEGER, 1989 Molecular cloning of cDNA encoding the p70 (Ku) lupus autoanitgen. J. Biol. Chem. **264**: 5047-5052.

RIO, D. C., 1990 Molecular mechanisms regulating Drosophila *P* element transposition. Ann. Rev. Genet. **24**: 543-578.

RIO, D. C., and G. M. RUBIN, 1988 Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element. Proc. Natl. Acad. Sci. USA **85**: 8929-8933.

ROBERTSON, H. M., and W. R. ENGELS, 1989 Modified *P* elements that mimic the *P* cytotype in *Drosophila melanogaster*. Genetics **123**: 815-824.

SEARLES L. L., A. L.GREENLEAF, W. E. KEMP and R. A. VOELKER, 1986 Sites of *P* element insertion and structures of *P* element deletions in the 5' region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6**: 3312-3319.

SHAPIRO, J. A., 1979 Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA **76**: 1933-1937.

SIEBEL, C. W., and D. C. RIO, 1990. Regulated splicing of the Drosophila *P* transposable element third intron *in vitro*: somatic repression. Science **248**: 1200-1208.

STRUHL, G., and K. BASLER, 1993 Organizing activity of wingless protein in Drosophila. Cell **72**: 527-540.

TAKASU-ISHIKAWA, E., M. YOSHIHARA and Y. HOTTA, 1992 Extra sequences found at *P* element excision sites in *Drosophila melanogaster*. Mol. Gen. Genet. **232**: 17-23.

TALBOT, W. S., E. A. SWYRYD, and D. S. HOGNESS, 1993 Drosophila tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. Cell **73**: 1323-1337.

TSUBOTA, S. I., and H. DANG-VU, 1991 Capture of flanking DNA by a *P* element in *Drosophila melanogaster*: creation of a transposable element. Proc. Natl. Acad. Sci. USA **88**: 693-697.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7**:1355-1363.

WILLIAMS, J. A., and S. B. CAROLL, 1993 The origin, patterning and evolution of insect appendages. Bioessays **15**: 567-578.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988a Molecular analysis of hybrid dygenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol. Cell. Biol. **8**: 1489-1497.

WILLIAMS, J. A., S. S. PAPPU and J. B. Bell, 1988b Suppressible *P*-element alleles of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **212**: 370-374.

WILLIAMS, J. A., A. L. AKTIN and J. B. BELL, 1990a The functional organization of the *vestigial* locus in *Drosophila melanogaster*. Mol. Gen. Genet. **221**: 8-16.

WILLIAMS, J. A., I. M. SCOTT, A. L. AKTIN, B. J. BROOK, M. A. RUSSELL and J. B. BELL, 1990b Genetic and molecular analysis of vg^U and vg^W: Two dominant *vg* alleles associated with gene fusions in Drosophila. Genetics **125**: 833-844.

WILLIAMS, J. A., J. B. BELL and S. B. CARROLL, 1991 Control of Drosophila wing and haltere development by the nuclear *vestigial* gene product. Genes Dev. **5**: 2481-2495.

WILLIAMS, J. A., S. W. PADDOCK and S. B. CARROLL, 1993 Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. Development **117**: 571-584.

WILLIAMS, J. A., S. W. PADDOCK, K. VORWERK and S. B. CARROLL, 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature **368**: 299-305.

YANEVA, M., J. WEN, A. AYALA and R. COOK, 1989 cDNA-derived amino acid sequence of the 86-kDa subunit of the Ku antigen. J. Biol. Chem **264**: 13407-13411.